University of California, Berkeley
U.C. Berkeley Division of Biostatistics Working Paper Series

Year 2009 *Paper* 255

Causal Inference in Epidemiological Studies with Strong Confounding

Kelly L. Moore* Romain S. Neugebauer†
Mark J. van der Laan‡ Ira B. Tager**

*Division of Biostatistics, School of Public Health, University of California, Berkeley, klmoore@stat.berkeley.edu

†Division of Research, Kaiser Permanente Northern California, Oakland, CA, Romain.S.Neugebauer@nsmtp.kp.org

‡Division of Biostatistics, School of Public Health, University of California, Berkeley, laan@berkeley.edu

**Division of Epidemiology, School of Public Health, University of California, Berkeley, ibt@berkeley.edu

http://biostats.bepress.com/ucbbiostat/paper255

Causal Inference in Epidemiological Studies with Strong Confounding

Kelly L. Moore, Romain S. Neugebauer, Mark J. van der Laan, and Ira B. Tager

Abstract

One of the identifiabilty assumptions of causal effects defined by marginal structural model (MSM) parameters is the experimental treatment assignment (ETA) assumption. Practical violations of this assumption frequently occur in data analysis, when certain exposures are rarely observed within some strata of the population. The inverse probability of treatment weighted (IPTW) estimator is particularly sensitive to violations of this assumption, however, we demonstrate that this is a problem for all estimators of causal effects. This is due to the fact that the ETA assumption is about information (or lack thereof) in the data. A new class of causal models, causal models for realistic individualized exposure rules (CMRIER), introduced in van der Laan and Petersen (2007), is based on dynamic interventions. CMRIER generalize MSM, and their parameters remain fully identifiable from the observed data, even when the ETA assumption is violated, if the dynamic interventions are set to be realistic. Examples of such realistic interventions are provided. We argue that causal effects defined by CMRIER may be more appropriate in many situations, particularly those with policy considerations. Through simulation studies, we examine the performance of the IPTW estimator of the CMRIER parameters in contrast to that of the MSM parameters. We also apply the methodology to a real data analysis in air pollution epidemiology to illustrate the interpretation of the causal effects defined by CMRIER.

1 Introduction

Marginal structural models (MSM) were introduced in Robins (1998) as a class of causal models for the investigation of the effect of a treatment or exposure on an outcome. They can be applied in air pollution epidemiology for the investigation of the effect of exposure to air pollutants on health outcomes, particularly in those studies whose results are likely to have policy relevance. More specifically in that context, they represent the distribution of potential outcomes that are defined based on air pollution interventions that result in identical pollutant levels experienced by all experimental units in the target population. We refer to such interventions as static interventions to compare them to dynamic interventions, which may not result necessarily in identical pollutant levels for all experimental units.

Ideally, fitting an MSM requires the design of an experiment where one could observe for each unit its outcome under each possible exposure level. Such an experiment cannot be conducted in practice and, instead, only one of the potential outcomes can be observed on each unit in a real life experiment (e.g., randomized or observational study). This explains why potential outcomes are also referred to as counterfactual outcomes. The hypothetical data from the ideal experiment are referred to as full data. When the effects of interest are defined by interventions on air pollution exposure at one point in time, the full data are denoted by $X = (W, (Y_a : a \in \mathcal{A}))$, where Y_a represents the counterfactual outcome that corresponds to pollutant level a; $\mathcal{A}$ represents the set of possible pollutant levels; and W is a vector of covariates measured before the exposure. This is opposed to the observed data from a real life experiment denoted by $O = (W, A, Y)$, where Y and A are the outcome and pollutant level actually observed. Note that this counterfactual framework and notation can easily be generalized to longitudinal data problems, where the effects of interest are defined based on interventions on a history of pollutant levels. For clarity only, the results presented in this paper are not illustrated with the more complex longitudinal data structures.

MSM are models for the full data distribution. Specifically, MSM represent the aggregate effects on the outcome Y caused by static interventions on the exposure A at the population-level or possibly conditional on some baseline covariate(s) (denoted with $V \subset W$) that define population subgroups of interest. Typically, in practice, MSM are models for $E(Y_a)$, i.e., the mean of the counterfactual outcomes Y_a. The coefficients of such MSM can thus be interpreted causally as representing the average population-level effect of the exposure on the outcome. If particular subgroups of the target population are of interest, then MSM are typically models for $E(Y_a \mid V)$, i.e., the mean of the counterfactual outcomes Y_a for each population stratum defined by V. Thus, the coefficients of such MSM represent the average effect of the exposure on the outcome within each subgroup defined by the baseline covariate(s) V, e.g., gender. The general notation for an MSM in this paper is $m(a, V \mid \beta)$ where β is the coefficient (a.k.a causal parameter of interest) for which investigators wish to draw inference. For instance, $m(a, V \mid \beta) = \beta_0 + \beta_1 a + \beta_2 V + \beta_3 aV$ is a linear MSM that can be used with continuous outcomes and $m(a, V \mid \beta) = 1/(1+\exp -((\beta_0 + \beta_1 a + \beta_2 V + \beta_3 aV))$ is a logistic MSM that can be used with dichotomous outcomes.

The fundamental problem of causal inference and, thus, of MSM estimation is that the ideal experiment cannot be implemented in practice; only a subset of the full data can be

collected in real life: the observed data. Therefore, causal inference based on MSM is a missing data problem. This statement is reflected by the so-called consistency assumption denoted with $Y = Y_A$, i.e., the actual outcome collected on any unit is one of the potential outcomes for that unit contained in the full data. Based on this assumption, the observed data $O = (W, A, Y_A)$ indeed can be viewed as a missing data structure on the full data $X = (W, (Y_a : a \in \mathcal{A}))$, with the exposure A being the missingness variable.

In practice, proper causal inference must rely on identifiability assumptions, i.e., assumptions that guarantee that there is enough information contained in the observed data to infer the MSM parameter defined from the full data. Two such assumptions have been emphasized (Robins et al. (2000); Neugebauer and van der Laan (2005); van der Laan and Robins (2003)) in the literature: the untestable sequential randomization assumption (SRA) and the testable experimental treatment assignment (ETA) assumption.

The first assumption also is known as the no unmeasured confounders assumption. It ensures that a causal association between the exposure and the outcome can be disentangled from any other potential "spurious" associations between the exposure and the outcome. By definition, such "spurious" associations are not explained by the effect of the exposure on the outcome. The information that allow such disentanglement can be referred to as measurements on confounders of the effect of the exposure on the outcome. The SRA states that the collection of baseline variables W contain all such confounders. Formally, the SRA can be defined based on the concept of the exposure mechanism. We denote the conditional probability distribution of a discrete exposure A given the full data with $g(a \mid X) \equiv P(A = a \mid X)$ and refer to it as the exposure mechanism. In our air pollution example, the SRA states that the pollutant level, A, experienced by any unit is conditionally independent of the unit's full data, X, given the unit's baseline characteristics, W: $g(A \mid X) = g(A \mid W)$. In words, the exposure to the pollutant, A, is randomized within strata defined by the potential confounders W of the target population.

The second assumption is also known as the positivity assumption. It requires that there is sufficient information in the observed data to separate the effect of the exposure from the effects of the aforementioned measured confounders by imposing that each subject in the target population can experience all possible pollutant levels, $\mathcal{A}$, no matter what values that the confounders take. Formally, the ETA is defined as,

$$\min_{a \in \mathcal{A}} g(a \mid W) > 0, \text{a.e.}$$

For example, consider the study described in Moore et al. (2008), which aims to estimate the causal effect of ozone concentration on the proportion of asthma-related hospital discharges. In this study, geographical grids are the independent experimental units (rather than the inhabitants in the grids). Suppose that no grids with average temperatures under 65 degrees are exposed to high ozone levels. Thus, the ozone level (above or below the standard) is set deterministically based on the grid's temperature. In such a situation, it is impossible to fully disentangle the effect of ozone exposure on the outcome from the effect of temperature in the target population without making additional untestable assumptions. This limitation is due to the violation of the ETA assumption.

Scenarios such as the one just described are common occurrences that can result in biased estimation of the MSM causal parameter, β. Four approaches to estimate MSM parameters have been proposed (Robins (1986, 1987, 2000); Robins and Rotnitzky (2001);

van der Laan and Robins (2003); van der Laan and Rubin (2006)): the G-computation, Inverse Probability of Treatment Weighted (IPTW), Augmented IPTW (A-IPTW), and Targeted Maximum Likelihood (TMLE) estimators.

The IPTW estimator has been the most widely applied estimator of MSM parameters in practice. When the ETA assumption is violated theoretically, the IPTW estimator is inconsistent (i.e., biased in practice). When the ETA assumption is violated practically, it suffers from important finite-sample bias and large variability. Intuitively, the IPTW estimator can be viewed as a statistical tool that properly reweights the observations in the observed data such that the resulting "ghost" data mimic the data that could have been collected from a randomized trial where confounding of the exposure effect on the outcome does not exist (Robins et al. (2000)). If certain exposures in the observed data never (or rarely) occurred for a subgroup defined by certain characteristics (e.g., high ozone concentration in grids with under-represented minorities), then, there is insufficient information to mimic the data that could have be obtained from a randomized trial, since reweighting can only modify the distribution of observations that are observed at least once. The poor finite-sample performance of the IPTW estimator has been demonstrated even in the presence of practical ETA violations in Neugebauer and van der Laan (2005), and also in Wang et al. (2006). In the latter reference, a diagnostic tool to evaluate the potential bias in the IPTW estimator due to ETA violations is provided. These practical violations occur commonly in observational studies where the investigator does not control exposure assignment but can also occur in randomized trials due to chance only, particularly in the circumstance when the sample size is small.

The G-computation, A-IPTW or TMLE estimators can consistently estimate the MSM parameters even when the ETA assumption is violated in theory, and, with good finite-sample performance when the ETA assumption is violated practically. However, their good properties in such scenarios depend heavily on artificial information from parametric modeling assumptions instead of real information in the data (van der Laan and Robins (2003); Neugebauer (2006)). Consider again the ozone study example where no grids with average temperatures under 65 degrees are exposed to high ozone levels. To estimate the effect of ozone on the proportion of asthma-related hospital discharges, one could assume a linear model in temperature and ozone. Extrapolation to the area of the curve that corresponds with low temperature and high ozone based on the model fit could then be used. However, since there are no data in this area of the curve, the extrapolated estimates rely entirely on the parametric modeling assumptions (i.e., the linear model one assumed for the association between ozone and temperature with the proportion of asthma-related hospital discharges). These parametric modeling assumptions cannot be tested from the data, and; therefore, it is desirable not to rely on such assumptions in practice to avoid drawing incorrect inferences and, in particular, to avoid over-confidence in the inference derived from these estimation procedures.

In the next section, we illustrate with a simulation study the difficulties in drawing reliable causal inference with MSM when the ETA assumption is violated. Several *ad hoc* methods have been proposed to overcome the poor practical performance of MSM estimation when the ETA assumption is violated. We discuss these approaches in section 3. Each of these attempts maintain the same objective: to estimate the effect of static interventions on the exposure of interest. An ETA violation should be regarded as an indication that

there are insufficient data available to support the proposed analysis(es); the proposed investigation is too ambitious and unrealistic at the time the study was conducted. In some instances one may even regard the investigation of such interventions as irrelevant if such interventions cannot be implemented with the current state of knowledge. For instance, we could argue that it is not possible (with the current air pollution regulatory policies in place or feasible in the future) to lower ambient ozone concentrations beyond a certain threshold concentration, on high temperature days, in the face of extant concentrations of transported pollutants. Thus, it would be irrelevant to investigate the effect of low exposure to ozone when temperature is high under such circumstances. Instead, one could try to answer a more realistic causal question that remains of interest, i.e., more realistic in that the question can be addressed with the data at hand without the need for additional and untestable parametric modeling assumptions. Such causal questions can be defined based on the concept of dynamic interventions on the exposure of interest that are deemed "realistic".

A class of causal models that generalize MSM has been developed independently by van der Laan (2006) (see also van der Laan and Petersen (2007)) and Orellana L (2006) (see also Robins et al. (2008)) and illustrated recently (Bembom and van der Laan (2007b)) to investigate such "realistic" causal questions. In section 4, we review this class of models referred to as causal models for realistic individualized exposure rules (CMRIER) and further illustrate their importance as they apply to air pollution epidemiology in section 5. Finally, we discuss the application of the results presented in this paper in section 6. First, however, we provide a simulation study in section 2 that underscores the practical limitations of causal inference with MSM but also with traditional regression methods in air pollution problems with extreme confounding. The motivational example presented in section 2 will be carried forward to illustrate the material presented in the subsequent sections and in particular in section 3 where we review some ad hoc methods that have been proposed to overcome the problem of practical or theoretical violation of the ETA assumption.

2 Motivational Example

2.1 Specific aim

To illustrate the difficulties in drawing inference about causal effects defined with static interventions when the ETA assumption is violated, we present the following example. In a simulation study, we examine the effect of a continuous exposure A on an outcome Y that is confounded by a single confounder W_1. The simulated data were constructed such that the exposure level experienced by any unit is almost deterministically set by the confounder level. The goal of the study is to illustrate the issues that arise in MSM estimation in the presence of ETA violations.

In addition, we also illustrate the failure of traditional estimation approaches in this setting. Traditionally, the causal effect of an exposure A on an outcome Y is investigated by regressing Y on A and the confounders, W, allowing for a flexible adjustment for W through model selection, e.g., squared and/or logarithmic functions of W. This traditional approach relies on the correct specification of the parametric model for $E(Y \mid A, W)$. Note

that in practice, no interaction terms between the exposure A and confounders W typically are considered in the model selection involved in this approach. Thus, in practice, the accuracy of the derived causal inference often relies on the assumption that none of the potential confounders W also are effect modifiers, i.e., interactions between A and W indeed can be omitted from the model for $E(Y \mid A, W)$.

Furthermore, it is important to note that investigators often are not aware of the estimation challenge that is posed by violations of the ETA assumption whether the estimation approach adopted is based on a traditional regression model or an MSM with G-computation estimation. This is in contrast to IPTW estimation of MSM parameters in which the violations of the ETA assumption become apparent from the inflated variance estimates due to extreme weighting (for more details see section 3.1.2). Through the simulation study presented in this section, we not only illustrate the poor practical performance in estimating causal parameters with MSM but also with traditional regression models.

2.2 Simulation protocol

The simulated data were generated according to the following distributions:

- $\begin{pmatrix} W_1 \\ W_2 \\ W_3 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 20 \\ 4 \\ 3 \end{pmatrix}, \Sigma\right), \Sigma = \begin{pmatrix} 25 & 0.8 & 0.7 \\ 0.8 & 3 & 0.3 \\ 0.7 & 0.3 & 1.5 \end{pmatrix}$
- $W_4 = 0.6(W_1 - 20)/5 + \sqrt{(}1 - 0.6^2) + \epsilon_1,\ \epsilon_1 \sim \mathcal{N}(0, 10)$
- $W_5 \sim \mathcal{U}(0, 3)$
- $P(W_6 = 0) = 0.5$ and $P(W_6 = 1) = 0.5$,
- $A = W_1 + \epsilon_2,\ \epsilon_2 \sim \mathcal{N}(0, 5)$,
- $Y = 4A + 5W_1 + \epsilon_3,\ \epsilon_3 \sim \mathcal{N}(0, 400)$,

where a) $\mathcal{N}(\mu, \Sigma)$ represents the (multivariate) normal distribution with mean μ and standard error (matrix of variance-covariance) Σ, b) $\mathcal{U}(a, b)$ represents the uniform distribution between a and b. We denote $W = (W_1, ..., W_6)$. In this simulation, the true causal effect on Y from static interventions on A can thus be represented with the following MSM for $E(Y_a)$: $100 + 4a$. The above data generating distributions were used to simulate 25000 datasets for sample sizes ranging from 100 to 1000.

We denote with $\rho_{X,Z}$ the correlation between X and Z. The simulation protocol just described results in the following correlation structure between the baseline covariates:

- $\rho_{W_1,W_2} = \frac{0.8}{\sqrt{25}\sqrt{3}} \approx 0.09$
- $\rho_{W_1,W_3} = \frac{0.7}{\sqrt{25}\sqrt{1.5}} \approx 0.12$
- $\rho_{W_1,W_4} = 0.6$
- $\rho_{W_1,W_5} = \rho_{W_1,W_6} = 0$

- $\rho_{A,W_1} = \frac{25}{\sqrt{50}\sqrt{25}} \approx 0.71$

The simulated variable W_1 is correlated with W_4 and is moderately correlated with W_2 and W_3. Similarly, W_1 is relatively highly correlated with the exposure of interest A. Thus, disentanglement of the effect of A on Y from the potential effects of W_1 on Y is a difficult problem in this simulation study due to the relatively high correlation between W_1 and some of the potential confounders but also between W_1 and A. We evaluated the performances of the IPTW and G-computation estimator of the MSM parameters and contrasted them to that of the traditional regression approach in this scenario.

Firstly, we applied the G-computation and IPTW estimators to fit the MSM, $m(a \mid \beta) = \beta_0 + \beta_1 a$, where the true value of the causal parameter defined by this MSM is $\beta = (\beta_0, \beta_1) = (100, 4)$.

- Implementation of the G-computation estimator is based on an estimate for $E(Y \mid A, W)$. The G-computation estimate is obtained from the estimate for $E(Y \mid A, W)$ by averaging out W through a second regression involving the MSM m (Robins (1986, 1987)). In this simulation study, three estimates for $E(Y \mid A, W)$ were used to derive three G-computation estimators. First, the estimate for $E(Y \mid A, W)$ was obtained with a stepwise (backward) model selection algorithm based on the Akaike Information Criterion (AIC) (stepAIC routine in R). The base model in this procedure is linear with a main term for A and all other potential confounders $(W_1, ..., W_6)$. Second, the same stepwise procedure was run without forcing A into the model. In the case that A was not selected by the algorithm, the G-computation estimate was thus 0. Third, the estimator for $E(Y \mid A, W)$ was obtained by fitting the correct linear model for $E(Y \mid A, W)$ known from the simulation protocol: $\alpha_0 + \alpha_1 A + \alpha_2 W_1$. Note that in practice this model is typically unknown.

- Implementation of the IPTW estimator is based on an estimate for the exposure mechanism, $g(A \mid W)$. In the simulation study, the estimate for $g(A \mid W)$ was always obtained based on the correctly specified model known from the simulation protocol: $\mathcal{N}(\gamma_0 + \gamma_1 W_1, \sigma)$. For further details on IPTW estimation, see section 3.1.2.

Secondly, we applied a traditional regression approach to estimate the effect of A on Y. This approach relies on an estimator for $E(Y \mid A, W)$. In the traditional approach, the coefficient for the main term of A is selected as the estimate of the effect of A on Y. Thus, no interaction terms between A and W can be considered when using this approach. If $E(Y \mid A, W)$ is estimated with a linear model with no interaction terms between A and W, then the estimate of coefficient in front of A in that model is always equal to the corresponding G-computation estimate of the MSM slope parameter β_1 based on this same estimate for $E(Y \mid A, W)$. For simplicity in demonstrating the results of this motivational simulation example, no interaction terms with A were considered in the stepwise model selection for $E(Y \mid A, W)$. Thus, the traditional and G-computation estimates are equivalent in this particular example. We note this equivalence does not hold generally.

2.3 Simulation results

The performance of the estimators is assessed based on two quantities: bias and mean squared error (MSE). The bias is estimated by computing the difference between the average

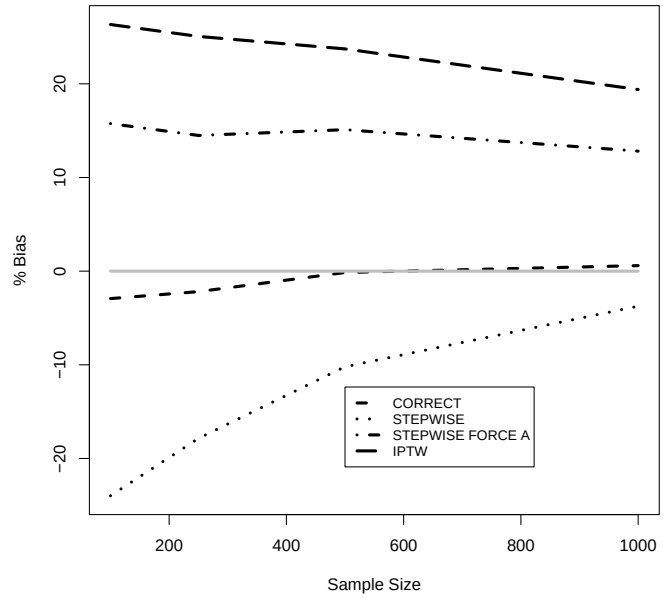
of the 25000 estimates and the true value at each sample size. The bias as a percent of the true value by sample size is provided in graph a) in Figure 1 for each estimator. We can immediately observe from this plot that the bias of the IPTW estimates is the largest of the four estimators across all sample sizes. Although less severe than the IPTW estimator, the bias of the traditional estimator based on the stepwise estimate of $E(Y \mid A, W)$ with A forced into the model is 12.8% at sample size 1000. The positive bias is a result of the correlation between A and W_1. Disentangling their individual effects on Y is difficult and thus the combined effect of A and W_1 is often attributed to A only when A is forced into the model for $E(Y \mid A, W)$. In contrast, the traditional estimator based on the stepwise estimate of $E(Y \mid A, W)$ without forcing A into the model is negatively biased. This is a result of incorrectly ascribing the combined effect of A and W_1 on Y largely to W_1. In practice, when the effect is entirely attributed to W_1, one would incorrectly conclude that there is no effect of the exposure A on the outcome Y. Finally, the traditional estimator (which is also the G-computation estimator) based on the correctly (in practice unknown) specified $E(Y \mid A, W)$ is consistent (bias $\approx 0\%$). We note the simulation was also run for a sample size of 50000 to verify that all estimators converged (bias $\approx 0\%$), however we have omitted those results for clarity of presentation.

The MSE can be decomposed as the sum of the squared bias of the given estimator and its variance. Therefore, the MSE was computed by summation of the squared bias, as explained above, and the sample variance of the 25000 estimates. For two estimators with similar bias, we can obtain a sense of their relative variance by comparing the MSE. Graph b) in Figure 1 provides the mean squared error (MSE) for the 25000 estimates for each of the four estimators. The comparison of the IPTW and G-computation estimates with respect to the MSE is striking. The relatively large MSE of the IPTW estimates is in part due to its larger bias but can also be attributed to its large variability. The IPTW estimator's poor finite sample performance is due to practical ETA violations. Since the exposure is assigned in a nearly deterministic fashion, for given values of W_1, the estimated probabilities of exposure are very low causing an inflation of the weights (see section 3.1.2 for details). Clearly the inference drawn from the IPTW estimator in this simulation study is unreliable. The MSE of the G-computation estimates based on the stepwise estimate of $E(Y \mid A, W)$ where A is forced into the model is comparable to the G-computation estimates based on the correctly specified $E(Y \mid A, W)$. Forcing A into the model for $E(Y \mid A, W)$ results in inflating the estimated effect of A on Y as explained above. This results in only a small amount of variability in that this inflated estimate is consistently estimated in the 25000 simulated datasets for each sample size. Thus, although this estimator is not variable, it is nonetheless biased. The G-computation estimator based on the stepwise selected $E(Y \mid A, W)$ without forcing A into the model is slightly more varible than the one the forces A into the model. Although it converges to the truth more quickly, again for any given dataset, the effect can be entirely attributed to either A or W_1 which results in a more variable estimator.

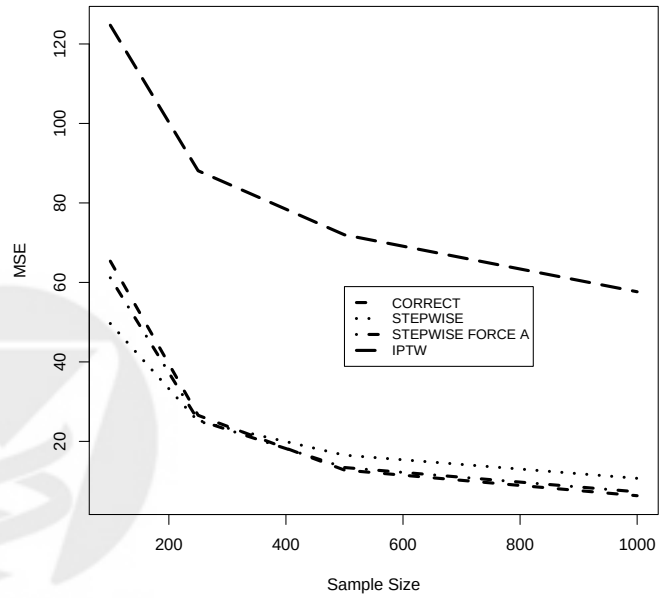
This simulation study illustrates the severe limitation from which the IPTW estimator suffers when the ETA assumption is violated. However, this issue is not exclusive to IPTW estimation of the MSM parameters, as as evidenced by the results from the G-computation and traditional estimation techniques. Even when the exposure is forced into the model, the effect estimator remains biased in finite sample sizes due to the fact that the data lack

information to enable identification of the effects of the exposure A versus confounders W on the outcome Y. With the G-computation and traditional estimation methods, we would often conclude incorrectly that there is no effect of the exposure on the outcome when the exposure is not forced into the model for $E(Y \mid A, W)$. In addition, if we ran this simulation with no effect of A on Y, these overly optimistic estimators (with respect to precision) could lead to false positive findings. Even though the IPTW estimator is also biased, due to its high variability, the danger of incorrectly concluding that there is an effect is not as high. Thus, in practice none of these estimators are reliable to answer the causal question of interest.

To overcome this issue, several methods have been proposed that we outline in section 3. Each of these methods (where applicable) is illustrated by applying them to these simulated data. We also apply the newly proposed CMRIER to an extension of these simulated data in section 5.



(a) Consistency



(b) MSE

Figure 1: Motivational Simulation Study: Consistency and MSE of IPTW and G-computation Estimates. "CORRECT" is the G-computation estimator, and equivalently in this simulation study the traditional estimator, when the model for $E(Y|A,W)$ is correctly specified; "STEPWISE" is the G-computation/traditional estimator when A is not forced into the selected model for $E(Y|A,W)$; "STEPWISE FORCE A" is the G-computation/traditional estimator when A is forced into the selected model for $E(Y|A,W)$; and "IPTW" is the IPTW estimator.

3 Causal effect estimation with MSM when the ETA assumption is violated

Several methods have been proposed to overcome the poor IPTW estimation performance illustrated above when the ETA assumption is violated. In the following subsections, we review these methods. The first subsection outlines three methodological proposals that attempt to improve estimation of the original MSM parameters from Robins (1998), described in the previous section. The next subsection outlines three methodological proposals that significantly differ in their approach to the problem, in that they focus on different MSM parameters for which violations of the ETA assumption are mitigated. In the last subsection, we discuss the practical importance and limitations of the different approaches proposed and motivate the application of CMRIER beyond or in combination with these approaches. Where applicable, we apply the specific method to the simulation study from section 2.

3.1 Methods bound to the original MSM parameter

3.1.1 Beyond IPTW estimation

Investigators can rely on the G-computation, A-IPTW or TMLE estimators for MSM estimation instead of the IPTW estimator in practice. The choice of the IPTW estimator often has been driven by its ease of implementation. However, it may be appealing to turn to a different estimator when a violation of the ETA assumption is detected in practice. In theory, and unlike the IPTW estimator, the other estimators indeed can maintain their consistency and efficiency properties, even when the ETA assumption is violated. However, it is important to realize that causal inferences drawn by these estimators in practice depend heavily on untestable parametric model assumptions instead of real causal information in the data when the ETA assumption is violated in theory or empirically (Yu and van der Laan (2002); Neugebauer and van der Laan (2003); Neugebauer (2006)). Indeed, all three estimators can artificially allow disentanglement of the effect of the exposure from the effects of confounders by relying on extrapolations from parametric models. In addition, another major concern with blind application of these estimators is the risk of not recognizing violations of the ETA assumption, since they will not provide signs of poor performance through inflated variability estimates unlike the IPTW estimator. We illustrated these points in section 2 with the application of the G-computation estimator.

3.1.2 IPTW estimation with weight truncation

As noted above, IPTW estimation can be implemented easily in practice. Its implementation is based on a two-step procedure. The first step involves the estimation of the exposure mechanism, $g(A \mid W)$ (under the SRA assumption). The second step involves a weighted regression of the outcome on the exposure of interest with weights inversely proportional to the exposure mechanism evaluated at the experienced exposure and confounder levels. The consistency of IPTW estimation – i.e., in practice its unbiasedness – relies on consistent estimation of the IPTW weights (i.e., the exposure mechanism), but also on the upholding of the ETA assumption as illustrated in section 2.

It is clear from the implementation of the IPTW estimator that some weights will be largely inflated when an exposure level is rarely observed in the data for some levels of the confounders, i.e., when the ETA assumption is so-called practically violated. Such inflated weights result in an MSM fit driven by a few observations in the data, which, in turn, leads to important finite sample bias and weak precision in the derived IPTW estimate (Neugebauer and van der Laan (2005)).

One approach that has been proposed to overcome this poor finite sample performance of the IPTW estimator is to truncate the weights, e.g., between 1.11 and 10 which corresponds with conditional probabilities of exposure given the confounder levels between 0.1 and 0.9 respectively (Cole and Hernan (2008)). This approach aims to limit the maximum contribution that any observation in the data can have on the MSM fit. A direct result of weight truncation is the reduction in the variability of the IPTW estimator. However, a side effect of weight truncation is the introduction of bias in the estimation of the weights that, in turn, leads to bias in IPTW estimation. This proposed approach can be viewed as balancing the bias-variance trade-off in IPTW estimation: the increase in the precision of IPTW estimation through weight truncation is at the cost of increased IPTW bias due to inconsistent estimation of the exposure mechanism. The performance of weight truncation can be assessed in practice by comparison of the MSE achieved by the truncated IPTW estimator and the (untruncated) IPTW estimator. Such comparisons have demonstrated that weight truncation, based on an *ad hoc* truncation level like the one suggested above, can actually lead to an increase in MSE (Bembom and van der Laan (2008b)). That is why it has been argued that the level of weight truncation should involve the sample size in practice, e.g., no observation should receive a weight greater than 10% of the sample size. To this end, Bembom and van der Laan (2008b) proposed a data-adaptive method for selection of the truncation level that achieves the optimal bias-variance trade-off, i.e., minimization of the MSE from the truncated IPTW estimator.

Finally, it is important to realize that weight truncation has been proposed to overcome the poor finite sample performance of IPTW estimation when the ETA assumption is only practically violated. Indeed, this approach is not suitable to address theoretical violations of the ETA assumption; the IPTW weights do not even suffer from the inflation described above when some exposure levels are not observed at all for certain levels of the confounders in the data.

For an illustration of this approach, we return to the simulation example provided in section 2 where the MSM was fit with the IPTW estimator with the correctly specified exposure model. For comparison purposes, the truncated IPTW estimator was applied to this simulation study. The truncation levels for the weights are 1/0.9=1.1 and 1/0.1=10. Graph a) in Figure 2 provides the consistency results for the IPTW and truncated IPTW estimators. Clearly, a large amount of bias was introduced by truncating the weights, by a factor of approximately 2 over the IPTW estimator with the correctly specified exposure mechanism. Furthermore, the truncated IPTW estimator is not converging towards the true value as the sample size increases, whereas, the IPTW estimator with the correctly specified exposure mechanism is in fact slowly converging. Note that the bias is reduced from 19% at a sample size of 1000 to 15% at a sample size of 50000. The MSE for the two estimators provided in graph b) in Figure 2 demonstrates the large reduction in variability as a result of truncating the weights. In comparison to the IPTW estimator with the

correctly specified exposure mechanism, the bias of the truncated IPTW estimator is more severe, whereas its variability is significantly lower, resulting in an overall substantially smaller MSE. This example clearly demonstrates the bias-variance trade-off that is a result of truncating the weights. In the case that there is truly no effect of A on Y, such a biased estimator with overly optimistic precision can lead to incorrect rejection of the null hypothesis of no effect.

3.1.3 IPTW estimation with curtailed adjustment for confounding

Another approach has been proposed to address extreme weighting with IPTW estimation. It is based on the exclusion of the confounder(s) that are causing weight inflation. In particular, if one confounder is highly predictive of the exposure but is known to have only a weak effect on the outcome or is weakly associated with it, then one may consider removing this covariate from the adjustment set of confounders used to estimate the exposure mechanism. Removing such confounders can result in an important increase in the precision of IPTW estimation at the cost of only a small increase in bias. Indeed, the confounder that is ignored can be deemed a weak confounder due to its weak effect on (or association with) the outcome. Ignoring weak confounders results in a limited violation of the aforementioned SRA assumption and, thus, negligible estimation bias compared to the large gain in estimation precision due to the weight stabilization induced by ignoring such confounders. Note that, in practice, and as a first approximation, identification of weak confounders can be based on subject-matter knowledge or based on the assessment of the crude univariate association between the confounder and the outcome. In Bembom and van der Laan (2008a), a formal method for selecting the adjustment set of confounders was proposed.

This approach can be illustrated in our study of the effect of ozone on the proportion of asthma-related hospital discharges (Moore et al. (2008)). Only those covariates with univariate association p-values less than 0.05 with ozone and the proportion of asthma-related hospital discharges were considered as candidates for the exposure model. However, we could have further reduced this set by excluding PM10 which had a p-value for the univariate association with ozone and proportion of asthma-related hospital discharges of $< 10^{-163}$ and 0.01 respectively. Under the assumption that PM10 is only a weak confounder, we could have removed it from the exposure model to potentially reduce the variability due to the extreme weights with the knowledge that a certain amount of bias (hopefully small) could result.

3.2 Methods that explore different MSM parameters

3.2.1 MSM with categorical exposures

Research questions that involve exposure(s) measured on a continuous scale also typically involve important practical violations of the ETA assumption. This is because of the very low probability of observing any given exposure level in each of the population subgroups defined by the levels of the confounders. In this situation, an approach that can be used by investigators consists of discretizing the exposure into two or more categories with the expectation that the ETA assumption associated with the new categorical exposure will no longer be violated or at least less violated. Beside mitigating the violation of the ETA

assumption, this approach also alters the definition of the causal effect of interest, i.e., new MSM parameters are defined based on the categorical exposure variables (see Appendix A.1).

For an illustration of this method, we return to the simulation study introduced in section 2. We first discretize the exposure into a three level categorical exposure variable from the original continuous exposure variable, at roughly the 33rd and 66th percentiles, defined by,

$$A^* = \begin{cases} 0 & \text{if } A < 15 \\ 1 & \text{if } 15 \leq A < 25 \\ 2 & \text{if } A \geq 25 \end{cases}$$

The observed data become $O = (W, A^*, Y)$. We now investigate the effect of A^* on Y with the MSM: $E(Y_{a^*})$: $m(a^* \mid \beta) = \beta_0 + \beta_1 I(a^* = 1) + \beta_2 I(a^* = 2)$. From the simulation protocol, we can derive the true value for the new MSM parameter $\beta \approx (148.2, 31.7, 63.5)$ (see Appendix A.1 for details on obtaining the true values of the new MSM parameter). Note that this approach is based on the definition of interventions on A corresponding with setting A^* to 0 and 1 (see Appendix for such definitions).

The MSM was fit with the IPTW estimator, based on the correctly specified exposure model for $P(A^* \mid W)$ (for details see Appendix A.1). Figure 3 provides the consistency results by plotting the bias of the 25000 IPTW estimates of β_0, β_1 and β_2 by sample size as a percent of the true value. Even at a sample size of 1000, the estimates of β_1 and β_2 have not yet converged to their true values. Since the exposure mechanism is correctly specified and convergence is still not achieved for at sample size 1000, evidently discretizing exposure into three categories did not adequately remove the ETA violations.

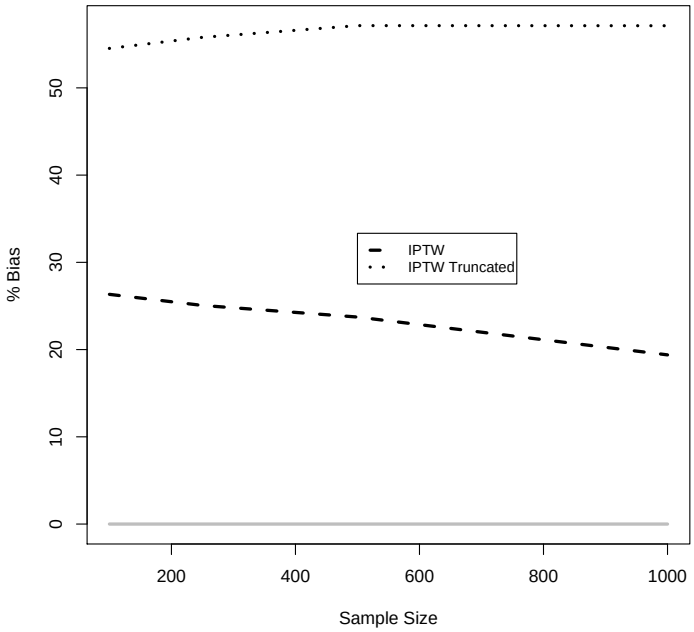
In a more aggressive attempt to mitigate the ETA violations, a binary variable A^* was created where A^* computed from A as $A^* = 1$, if $A > \theta, \theta = 20$, and $A^* = 0$ if $A \leq \theta$. Note that the cutoff value of 20 is near the median of A. We now investigate the effect of A^* on Y with the MSM: $m(a^* \mid \beta) = \beta_0 + \beta_1 a^*$. We can again derive the true value for the new MSM parameter β as described in Appendix A.1: $\beta \approx (161.9, 36.2)$.

The MSM was again fit with the IPTW approach, also based on the correctly specified exposure model (for details see Appendix A.1).

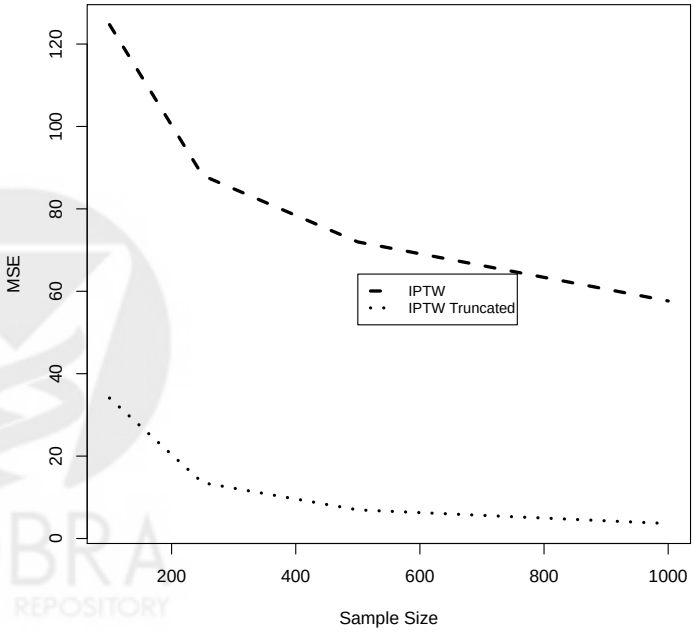
Figure 4 provides the consistency results for the IPTW estimator of the intercept (β_0) and slope (β_1) parameters from the MSM with the binary exposure. With a comparison of Figure 4 to Figure 3, we can observe visually that the biases of the IPTW estimates of β_0 and β_1 for the binary exposure are smaller than those of β_0, β_1 and β_2 for the categorical exposure. The average bias for the IPTW estimates of β_0 and β_1 for the binary exposure is approximately 3% as compared to the average bias of the IPTW estimates of β_0, β_1 and β_2 for the categorical exposure of 8%. This indicates that reducing the number of categories for the exposure variable reduces the ETA violations.

This simulation demonstrates that one can potentially mitigate the ETA issue by discretizing a continuous exposure variable, given an adequate sample size. However, this is at the expense of significant alteration of the causal question of interest. Thus, this approach should only be applied if the new causal question is of interest from a subject matter perspective. For instance, this approach is reasonable in air pollution epidemiology to test the effect of air pollutants, e.g., dichotomized pollutant levels based on the current regulatory standard (low versus high). One can more closely approximate the original effect of interest

by creating a number of categories, however this of course increases the potential for ETA violations.



(a) Consistency



(b) MSE

Figure 2: Simulation Study: IPTW and IPTW with truncated weights

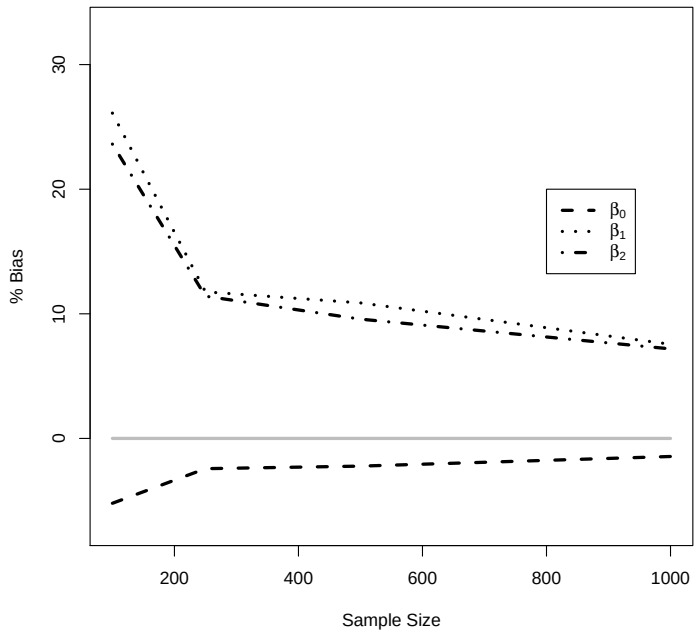


Figure 3: Simulation Study with discretized categorical exposure: Consistency

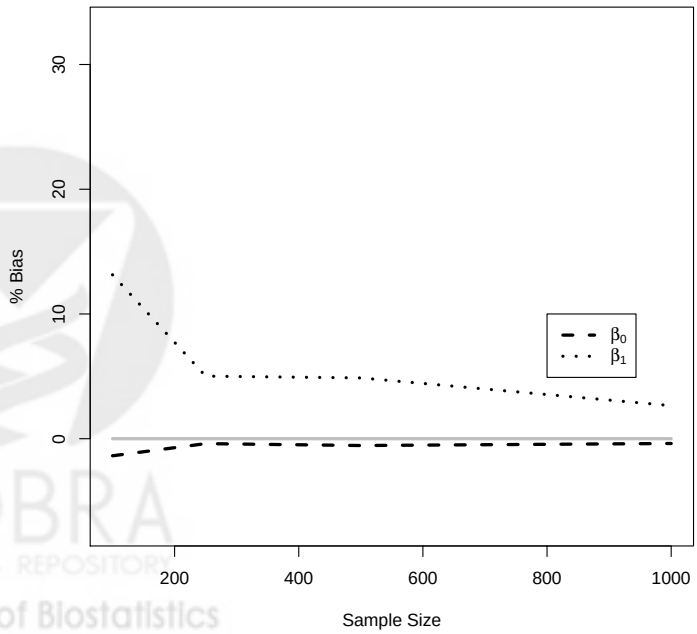


Figure 4: Simulation Study with discretized binary exposure: Consistency

3.2.2 MSM for data-driven population subgroup(s)

Filtering out from an MSM analysis the observations with no exposure experimentation, i.e., those that violate the ETA assumption, is a natural approach to overcome the poor finite sample performance of the IPTW estimator. Indeed, if one removes the observations for which the levels of the confounders deterministically limit the set of exposure levels that can be experienced by an experimental unit, then, the IPTW weights associated with the remaining observations in the data do not suffer from abnormal inflation. While this approach is appealing, it is important to realize that such a filtering of the data amounts to implicit modification of the scope of the initial research question by focusing on a subgroup of the initial target population. In other words, a new MSM parameter becomes the parameter of interest; and the relevance of this parameter to address the subject-matter research question of interest is not guaranteed.

Indeed, the exposure mechanism is based on a number of confounders in practice; and violation of the ETA assumption is typically not caused by one such confounder only but a combination of these confounders. The population subgroup that results from filtering out the observations that violate the ETA assumption cannot easily be interpreted necessarily since it is composed of observations characterized by complex combinations of confounder levels. While these complex combinations justify their inclusion in the analysis, they likely do not lead to the definition of an interesting population subgroup for the purpose of defining a relevant exposure effect for the subject-matter research of interest. In particular, this approach cannot be applied successfully to research questions that involve time-dependent confounders (longitudinal data where cumulative effects over time are of interest) that cause ETA violations and that are also on one of the causal pathways of interest (Robins et al. (2000)). In general, this approach will thus not apply to problems that involve the investigation of the effect of a history of exposure (i.e., experienced over more than one time point) on an outcome.

Thus, the application of this approach is limited in practice to isolated cases where the ETA assumption is clearly violated due to only one or two baseline covariates which define interesting population subgroups of interest. In such situations, instead of removing the observations with no exposure experimentation, one may instead rely on a stratified MSM analysis conditional on the one or two covariates identified as causing ETA violation. Alternatively, one may rely on a pooled MSM analysis conditional on these covariates with the implementation of the stabilized IPTW estimator proposed by Robins et al. (2000). Note that whichever analytic strategy is adopted (exclusion of observations or stratification), the sample size of each analysis involved is reduced. Such a sample size reduction may in turn result in new practical violations of the ETA assumptions (since such violations are a function of the sample size).

3.2.3 History-restricted MSM

This approach pertains to problems with longitudinal data which involve the investigation of the effect of a history of exposure (i.e., experienced over more than one time point) on an outcome. History-restricted MSM (HRMSM) for longitudinal data (Joffe et al. (2001, 2004); Neugebauer et al. (2007)), allow estimation of the causal effect of an exposure on an outcome based on a curtailed portion of the exposure history rather than the entire

exposure history of interest as is the case with standard MSM applications. By restricting the scope of the exposure history that defines the effect of interest, one mitigates practical violations of the ETA assumptions as noted by Joffe et al. (2001).

In our ozone data analysis, the effect of the history of quarterly ozone levels over 20 years on quarterly proportions of asthma-related hospital discharges during the same time span is of interest. In particular, one is interested in investigating the effect of ozone on the first quarter of the second year monitored. It is not reasonable to assume that ozone levels from the previous year would have a significant effect on the current quarter's proportion of asthma-related hospital discharges. In fact in Moore et al. (2008), it was reasoned that the effect of only the current quarter exposure to ozone on the same quarter proportion of asthma-related hospital discharges was sensible. Reducing the exposure history of interest to the latest experienced quarterly exposure is thus both sensible from a subject-matter point of view, but can also reduce the extent of the violation of the ETA assumption.

3.3 Beyond these proposed approaches

As discussed in the introduction, the violation of the ETA assumption can be viewed as a lack of information in the data to fully disentangle the exposure effect of interest from that of the confounders of this effect. Two fundamentally different paths can then be taken to address the lack of information in the data: 1) maintain the complexity of the initial research question and either a) heavily rely on parametric modeling assumptions that may not hold to draw causal inferences (e.g., rely on the G-computation estimator) or b) tolerate estimation bias (e.g., weight truncation); or 2) simplify the research question and draw causal inferences that can be supported fully by the information in the data (e.g., discretize exposure into a categorical variable).

We have described several methodologies that illustrate both of these paths. Among them, subject-matter investigators are most often dissatisfied with the methodology that consists in discretizing the exposure of interest due to the resulting simplification in the causal inference that can be attained with this methodology. Nevertheless, one can argue that such a simplification is typically necessary in practice because the information contained in finite sample data is often too limited to reliably (i.e., with a minimum of untestable assumptions) investigate effects defined based on continuous exposures. We believe that while the first path can provide exploratory information about subject-matter research problems on which future research can be based, it does not constitute a strong scientific basis for decision making (e.g., for regulatory/intervention purpose like setting air pollution standards). The second path should then be considered.

We described in section 3.2 three proposed methodologies that follow the second path described above. While some of these approaches can lead to reliable causal inference in practice, they typically need to be combined to successfully achieve reliable inference. However, even when these approaches are combined, the information in the data may still remain too limited to allow proper estimation of the MSM parameter of interest, i.e., the ETA assumption remains violated. This limitation motivates further exploration of the definition of less ambitious causal parameters that, nevertheless, can provide relevant causal inference for decision making. Such exploration lead to the definition of CMRIER introduced in van der Laan and Petersen (2007). This class of models builds on the coun-

terfactual framework on which is based the definition of MSM and constitutes, in fact, a generalization of MSM. We introduce them in the next section.

Before doing so, we would like to emphasize that, while the CMRIER parameters can be seen as less ambitious as MSM parameters when the ETA assumption is violated, they are, in fact, a class of causal models that can be used in situations where the ETA assumption holds. In other words, their use should not only be considered in cases where the ETA assumption is violated. Indeed, these models not only allow the investigation of causal effects defined by static interventions but also the effect of dynamic interventions that do not result in identical exposure levels for all experimental units. In fact, the effect of such dynamic interventions may be more interesting for some subject-matter research than the effect of static interventions (e.g., to optimize drug treatments) (see Bembom and van der Laan (2007a,b) for examples).

4 Causal effect models for realistic individualized exposure rules

As noted before, MSM are models which represent the effect of static interventions on an exposure of interest. A static intervention is one particular type of exposure intervention which results in a single exposure level a that is experienced by all units in the target population. However, an exposure intervention does not necessarily have to result in identical exposure levels for all units in the target population. Such interventions are deemed dynamic, since they may result in an exposure level experienced by a given unit that depends on that unit's characteristics W, i.e., the same dynamic intervention applied to two units could result in two different exposure levels experienced by each unit. We denote such a dynamic intervention with d. These interventions are a mapping from the unit's characteristics W to one of the possible exposure levels $d(W) \in \mathcal{A}$; therefore, they are also referred to as exposure rules. The effects of such dynamic interventions can be represented with a class of causal models, CMRIER. They are models for the distribution of the counterfactual outcome $Y_{d(W)}$. Note that one particular dynamic intervention is a static intervention, i.e., $d(W) = a$, and CMRIER can thus be viewed as a generalization of MSM.

When the ETA assumption is violated, we have argued that MSM parameters cannot be identified without making untestable assumptions due to the lack of true causal information in the data. This is because static interventions are not realistic interventions in the sense that the exposure levels resulting from these interventions are never seen in the collected data for some strata of the population defined by certain confounder levels.

Similarly, the estimation of CMRIER parameters also relies on an assumption analogous to the ETA assumption for MSM. This assumption requires that all units can experience the exposure levels resulting from the dynamic interventions, $d(W)$, for all values for the confounding variables that characterize each unit, W. We refer to such dynamic interventions as realistic dynamic interventions (or realistic exposure rules) to distinguish them from dynamic interventions (exposure rules) for which the ETA assumption is violated. Unlike MSM where investigators are limited to the investigation of effects from static interventions, CMRIER allow the investigation of effects from dynamic interventions that can be defined by the investigators themselves, so that the exposure rules considered are

always realistic and meaningful in the context of the subject matter. This control over the definition of causal effects is particularly appealing to overcome the limitation of the application of MSM when the ETA assumption is violated. Indeed, the investigators can focus their efforts on the investigation of causal effects that remain of interest but that also are fully identifiable from the data without additional untestable modeling assumptions like the ones that would be required to fit an MSM. Below, we discuss the interpretation of causal effects represented by CMRIER applied to the ozone data analysis, but first, we describe one method that investigators can use to define a dynamic intervention that can be deemed realistic based on any given static intervention from an MSM.

4.1 Realistic dynamic interventions

Consider a binary exposure variable, e.g., low versus high concentration of ambient ozone exposure. A dynamic intervention can be deemed realistic for a given unit only if the estimated probability of the exposure which results from that intervention, given the unit's confounder history, is greater than some threshold not too close to 0, e.g., 0.1. We denote such threshold value with α and later develop the practical meaning of our statement "not too close". The simplest and most natural dynamic interventions that can be considered are static interventions, e.g., exposure of every unit to an exposure level a that is independent of the confounder history for each unit. By definition of the violation of the ETA assumption, such static interventions are not realistic when the ETA assumption is violated. Instead, the investigators can then define a dynamic intervention that is realistic as follows: a unit is exposed to level a (0 or 1) if the probability of the unit's being exposed to level a given the unit's confounder history is greater than α and otherwise the unit is exposed to level $1 - a$. Formally, the proposed dynamic interventions above for a binary exposure A, are denoted with:

$$d(1)(W) = \begin{cases} 1 & \text{if } P(A = 1 \mid W) > \alpha \\ 0 & \text{otherwise} \end{cases}$$

$$d(0)(W) = \begin{cases} 0 & \text{if } P(A = 0 \mid W) > \alpha \\ 1 & \text{otherwise.} \end{cases}$$

We refer to $d(a)$ for $a = 0$ or $a = 1$ as exposure rules, because they are mapping from the unit's characteristics W into an exposure level $d(a)(W)$. The definition of such realistic dynamic interventions can be generalized easily to categorical exposures. For instance, consider an exposure with three ordered levels: 0, 1 and 2. Each of these three static interventions can be mapped into the following realistic dynamic interventions:

$$d(0)(W) = \begin{cases} 0 & \text{if } P(A = 0 \mid W) > \alpha \\ 1 & \text{if } P(A = 0 \mid W) < \alpha \text{ and } P(A = 1 \mid W) > P(A = 2 \mid W) > \alpha \\ 2 & \text{if } P(A = 0 \mid W) < \alpha \text{ and } P(A = 2 \mid W) > P(A = 1 \mid W) > \alpha \end{cases}$$

$$d(1)(W) = \begin{cases} 1 & \text{if } P(A = 1 \mid W) > \alpha \\ 0 & \text{if } P(A = 1 \mid W) < \alpha \text{ and } P(A = 0 \mid W) > P(A = 2 \mid W) > \alpha \\ 2 & \text{if } P(A = 1 \mid W) < \alpha \text{ and } P(A = 2 \mid W) > P(A = 0 \mid W) > \alpha \end{cases}$$

$$d(2)(W) = \begin{cases} 2 & \text{if } P(A=2 \mid W) > \alpha \\ 0 & \text{if } P(A=2 \mid W) < \alpha \text{ and } P(A=0 \mid W) > P(A=1 \mid W) > \alpha \\ 1 & \text{if } P(A=2 \mid W) < \alpha \text{ and } P(A=1 \mid W) > P(A=0 \mid W) > \alpha. \end{cases}$$

Alternatively, each of the three static interventions a can be mapped into realistic dynamic interventions where the alternative exposure level assigned when $g(a \mid W) < \alpha$ is not the most likely exposure levels based on $g(A \mid W)$ as described above, but the nearest exposure level among all exposure levels a that are also realistic, i.e., such that $P(A = a \mid W) > \alpha$:

$$d(0)(W) = \begin{cases} 0 & \text{if } P(A=0 \mid W) > \alpha \\ \min\{a : P(A=a \mid W) > \alpha\} & \text{if } P(A=0 \mid W) < \alpha \end{cases}$$

$$d(1)(W) = \begin{cases} 1 & \text{if } P(A=1 \mid W) > \alpha \\ \min\{\mid a-1 \mid : P(A=a \mid W) > \alpha\} & \text{if } P(A=1 \mid W) < \alpha \end{cases}$$

$$d(2)(W) = \begin{cases} 2 & \text{if } P(A=2 \mid W) > \alpha \\ \min\{\mid a-2 \mid : P(A=a \mid W) > \alpha\} & \text{if } P(A=2 \mid W) < \alpha. \end{cases}$$

The realistic exposure rules defined above can be used to define causal effects of the exposure of interest on an outcome through the specification of an CMRIER for $E(Y_{d(a)(W)} \mid V)$ denoted with $m(a, V \mid \beta)$. Note, that like MSM, CMRIER that are defined based on the realistic exposure rules described above are also functions of an exposure level a and the baseline covariate V that define population subgroups of interest. However, the causal interpretation of CMRIER is different from that of MSM in general and causal effects defined by such CMRIER may appear, at first, as irrelevant for the investigation of the original research question of interest. Therefore, we motivate their use in practice by illustrating their application and interpretation with the ozone study in the next section.

4.2 Causal interpretation

Returning to our ozone exposure data analysis, assume again that no grid with average temperatures below 65 degrees experienced ozone concentrations above the standard (i.e., level 1 of ozone). For these grids, the static intervention " set ozone above the standard" would not fall in the set of possible exposures. The static intervention of setting ozone to level 1 thus is not deemed to be realistic. One might be aware of such a violation of the ETA assumption in practice *a priori*. Most often, however, ETA violation is not discovered until the data are collected, as is the case in the ozone study. Certain exposure levels may be rare or not represented at all in the collected data for an unexpected subset of the population defined by a particular combination of confounder levels, e.g., low temperature levels. Whether the violation of the ETA assumption associated with static interventions is known *a priori* or discovered *a posteriori*, we can always define realistic dynamic interventions that map each observed confounder history into an exposure level as described in the previous section. In the ozone data analysis, the following two exposure rules can be used to define a causal effect of ozone on the proportion of asthma-related hospital discharges with an CMRIER:

- rule 1, $d(1)$: if the probability of experiencing an ozone concentration above the standard, given a unit's confounder history is less than α (i.e., the ETA assumption is violated at the α level); then, the exposure level should be set to 0, i.e., below the standard and otherwise to 1, i.e., above the standard.

- rule 2, $d(0)$: if the probability of experiencing an ozone concentration below the standard, given a unit's confounder history, is less than α (i.e., the ETA assumption is violated at the α level); then, the exposure level should be set to 1, i.e., above the standard and otherwise to 0, i.e., below the standard.

Investigation of the effect of ozone on the proportion of asthma-related hospital discharges can be based on the CMRIER associated with the exposure rules above: $E(Y_{d(a)(W)}) = \beta_0' + \beta_1' a$ to overcome the limitation of the application of the MSM: $E(Y_a) = \beta_0 + \beta_1 a$ due to violation of the ETA assumption. The interpretation of the causal parameter defined by the CMRIER can easily be understood by analogy with the interpretation of the MSM parameter. The MSM parameter β_1 in the ozone study represents the causal effect of ozone on the proportion of asthma-related hospital discharges by comparing: a) the mean proportion of asthma-related hospital discharges over all grids had all grids been exposed to an ozone level above the current standard (static intervention 1), $E(Y_1)$, to b) the mean proportion of asthma-related hospital discharges over all grids had all grids been exposed to an ozone level below the current standard (static intervention 2), $E(Y_0)$. Similarly, an CMRIER parameter β_1' represents the causal effect of ozone on the proportion of asthma-related hospital discharges by comparing the mean proportion of asthma-related hospital discharges under two different dynamic interventions: $E(Y_{d(1)(W)}) - E(Y_{d(0)(W)})$. A difference in the two means of proportions of asthma-related hospital discharges is interpreted as resulting from the interventions considered on ozone. This comparison under two different interventions is more relevant to policy considerations than the static intervention, since it restricts the intervention to units where such an intervention is possible. The CMRIER allows the approximation of the original causal effect defined by the MSM and, like MSM parameters, CMRIER parameters remain interpretable at the population level. Unlike MSM, however, CMRIER parameters remain fully identifiable from the observed data even when the ETA assumption is violated as long as the dynamic interventions on which the CMRIER are based are realistic such as the ones we proposed above.

In addition, it should be clear from the example above that CMRIER are a generalization of MSM. Indeed, if the ETA assumption was not violated in the ozone study, the realistic dynamic interventions above would correspond to the static intervention on ozone and the CMRIER would thus reduce to the standard MSM. Note, that when the set of possible exposure levels is limited to two exposure levels, then, the causal contrast, β_1 represented by an CMRIER converges to zero as the ETA assumption becomes more violated, i.e., it becomes more and more difficult to identify an effect due to interventions on the exposure of interest.

The methods used to estimate MSM parameters can be generalized to the estimation of CMRIER parameters. For the development of the estimation of CMRIER parameters see van der Laan and Petersen (2007). The details on the implementation of the IPTW estimator, see Appendix A.2. In addition, note that definition of causal effects in the ozone study above based on dynamic interventions on the dichotomized ozone exposures is based

on an implicit definition of the interventions "set $A^* = a^*$". Formal definition of such interventions is provided in Appendix A.2.

5 Applications

We provide two applications of the CMRIER methodology presented in section 4 through: 1) simulation studies, and 2) data analysis. The simulation studies are continuations of the simulation study introduced in section 2. The data analysis in which we demonstrate the application of the CMRIER methodology is the study that has been discussed throughout this paper which aims to investigate the effect of reductions of ozone concentrations on the proportion of asthma-related hospital discharges (Moore et al. (2008)).

5.1 Illustration of the CMRIER methodology with simulation studies

Data were again simulated according to the data generating distribution provided in section 2. To demonstrate the CMRIER methodology described above, a categorical exposure variable was created from the continuous exposure variable. The two discretization schemes described in section 3.2.1 were applied to obtain two exposure variables: the first binary and second categorical with three exposure levels. The effect of the new exposure variable A^* on the outcome Y was investigated by using both an MSM and an CMRIER (see Appendix for a detailed definition of this effect).

The parameters of both the CMRIER and MSM are estimated with the the IPTW estimation procedure. In both instances, the correctly specified exposure mechanism was applied (for details see Appendix A.1).

5.1.1 Binary exposure

The CMRIER for the binary exposure variable A^* is given by,

$$E(Y_{d(a^*)(W)}) = \beta_0^{'} + \beta_1^{'} a^*,$$

where $d(a^*)(W)$ is the rule presented in section 4 applied to A^* with $\alpha = 0.1$. The MSM is given by,

$$E(Y_a) = \beta_0 + \beta_1 a^*.$$

The true parameter values for the CMRIER coefficients are given by $\beta^{'} \approx (166.4, 27.1)$, and for the MSM coefficients they are given by, $\beta \approx (161.9, 36.2)$. Since the data were simulated such that the ETA assumption is violated, the parameter values are not equal. In this simulation, $\beta_1^{'}$ is approximately 25% lower than β_1. The CMRIER parameter is typically shrunk towards zero as compared to the corresponding MSM parameter. If there exist units in the data for whom an intervention is not possible, as defined by $g_n(a \mid W) < \alpha$, then there is no information in the data to estimate the effect for these units. The contribution of these units that are never exposed (to a given level) to the overall population-level effect defined by the two dynamic interventions is zero. Thus, the parameter of the CMRIER with a binary exposure is conservative (closer to zero) as compared to the MSM parameter.

The consistency results of the IPTW estimators of the CMRIER and MSM parameters are provided in graphs a) and b) in Figure 5. The IPTW estimator of the CMRIER is largely outperforming that of the MSM with respect to bias. Across all sample sizes, the bias of the IPTW estimator of the CMRIER parameter is substantially lower than that of the IPTW estimator of the MSM parameter with a maximum bias for the former of 3% in comparison to 13% for the latter.

Not only is the IPTW estimator of the CMRIER parameter outperforming the IPTW estimator of the MSM parameter with respect to bias, but also with respect to MSE. The MSE of the IPTW estimator of the MSM parameter is as high as 3 times that of the IPTW estimator of the CMRIER (Figure 6). Thus clearly the parameters of the CMRIER can be consistently estimated with far less uncertainty using the IPTW method even when the ETA is violated as it is in this example with an average across all simulated datasets of 20% (and median 20%) of the units meeting the criterion $g_n(A \mid W) < \alpha$ or $g_n(A \mid W) > (1-\alpha)$.

5.1.2 Categorical Exposure

To more closely approximate the original effect of interest, that is the effect of the continuous exposure A on Y, the exposure was also discretized into three levels. With a categorical exposure variable, more choices for the realistic intervention rule are available than the simple rule for binary exposures as described in section 4.1.

In this simulation study, we apply the rule that sets the realistic intervention to the corresponding static intervention when the probability of experiencing that intervention level is greater than α, otherwise it is set to the intervention that is most likely based on $g(A \mid W)$.

The CMRIER for the categorical exposure variable A^* is given by,

$$E(Y_{d(a^*)(W)}) = \beta_0^{'} + \beta_1^{'} I(a^* = 1) + \beta_2^{'} I(a^* = 2),$$

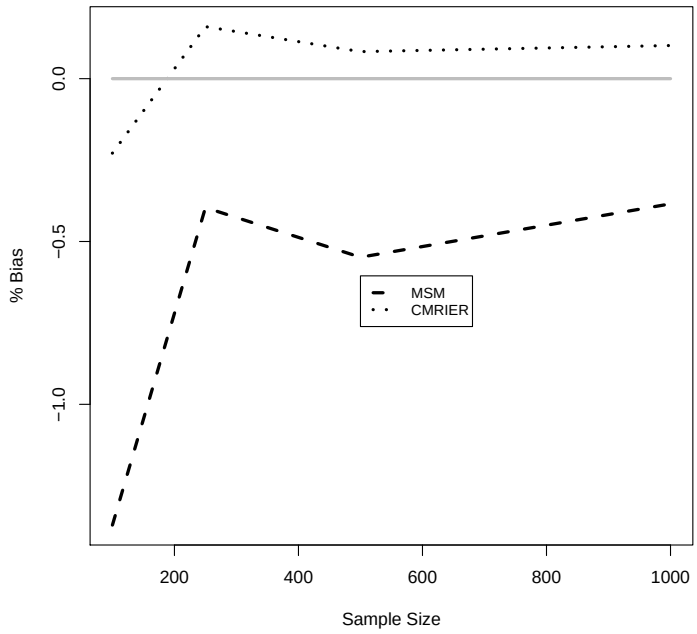
where $d(a^*)(W)$ is the rule applied to A^* with $\alpha = 0.1$. The MSM is given by,

$$E(Y_{a^*}) = \beta_0 + \beta_1 I(a^* = 1) + \beta_2 I(a^* = 2).$$

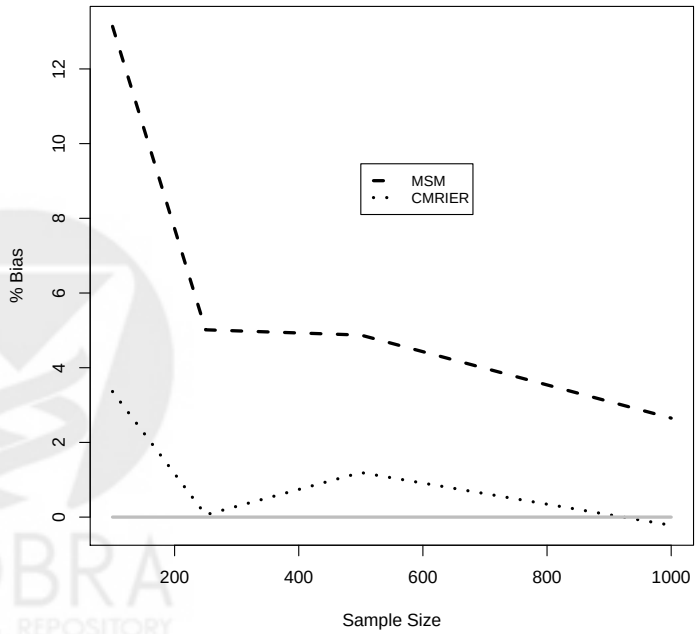
The true values for the CMRIER parameters are given by, $\beta^{'} = (148.2, 31.7, 63.5)$, and for the MSM parameters they are given $\beta = (161.6, 18.4, 36.9)$ (see Appendix A.2 for details on how these values were derived).

The consistency results for the IPTW estimator of the CMRIER and MSM parameters are provided in graphs a), b) and c) in Figure 7. The results are similar to those of the binary exposure in that the IPTW estimator of the CMRIER is largely outperforming that of the MSM with respect to bias. The bias is of the IPTW estimators of the CMRIER parameters is close to zero across all sample sizes and all parameters; the maximum absolute value of the bias for any of the parameters β_0, β_1 and β_2 is 2% at a sample size of 100, and decreases further as the sample size increases. In comparison, the bias for the IPTW estimator of the MSM parameter is much higher, with an absolute value of bias of 26% for β_1 at a sample size of 100, and, even at a sample size of 1000 the bias of the estimator of this parameter is 8%.

The MSE for the IPTW estimators of the CMRIER and MSM parameters are provided in graphs a), b) and c) in Figure 8. The comparison of the MSE indicates that in addition to the reduction in bias, there is also a large reduction in variability as in the binary simulation example.

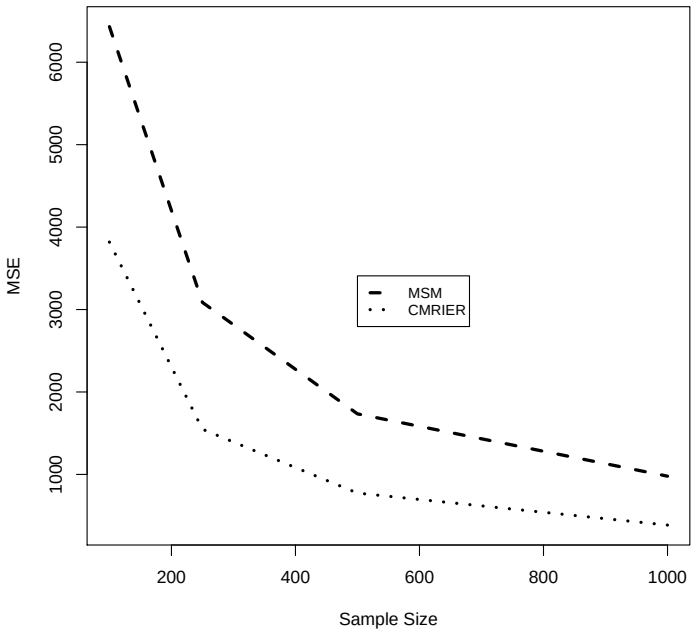


(a) β_0 and $\beta_0^{'}$

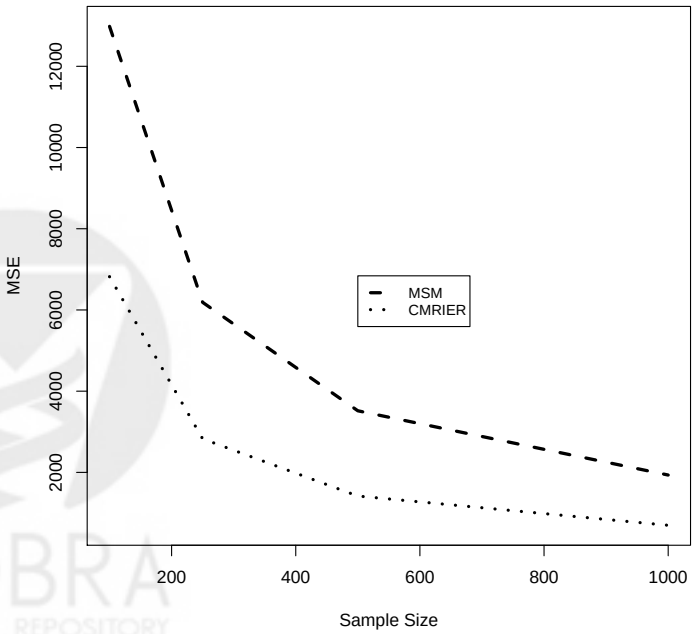


(b) β_1 and $\beta_1^{'}$

Figure 5: Simulation study with binary exposure: IPTW estimator of MSM and CMRIER parameters, consistency

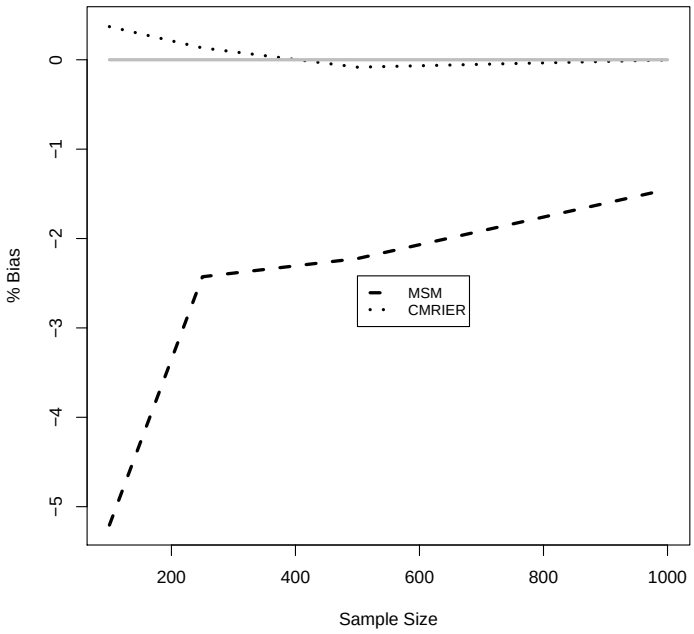


(a) β_0 and $\beta_0^{'}$

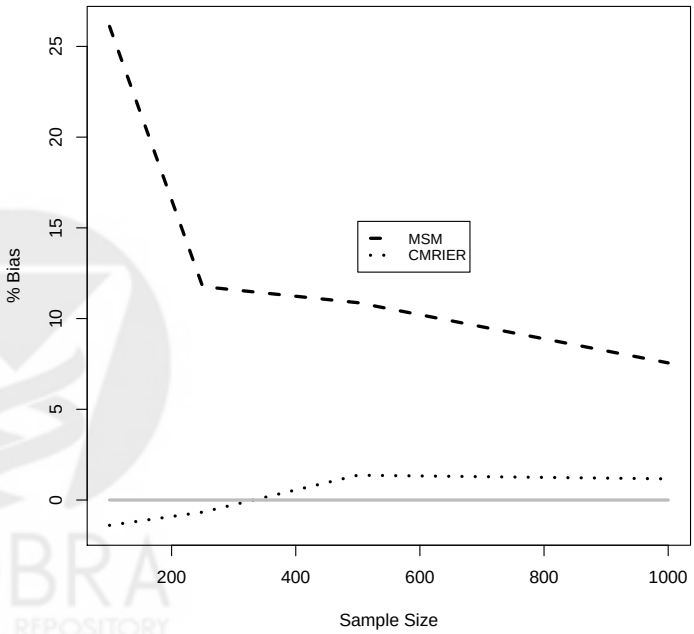


(b) β_1 and $\beta_1^{'}$

Figure 6: Simulation study with binary exposure: IPTW estimator of MSM and CMRIER parameters, MSE

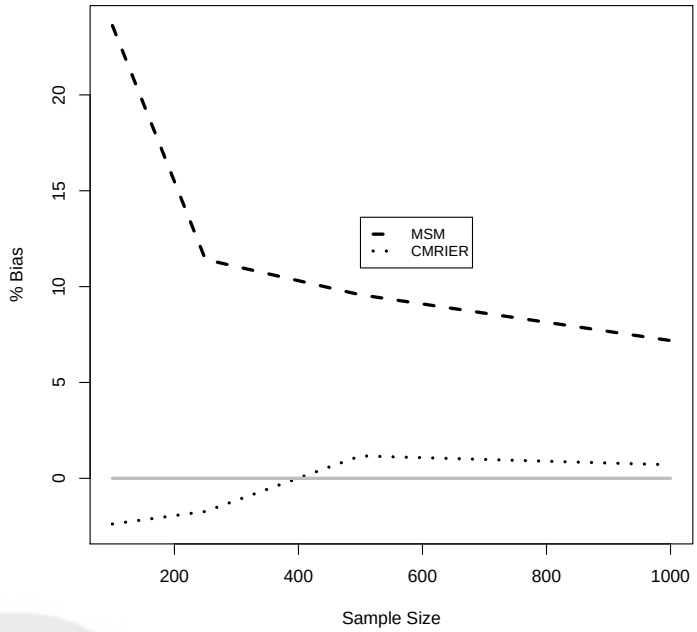


(a) β_0 and $\beta_0^{'}$



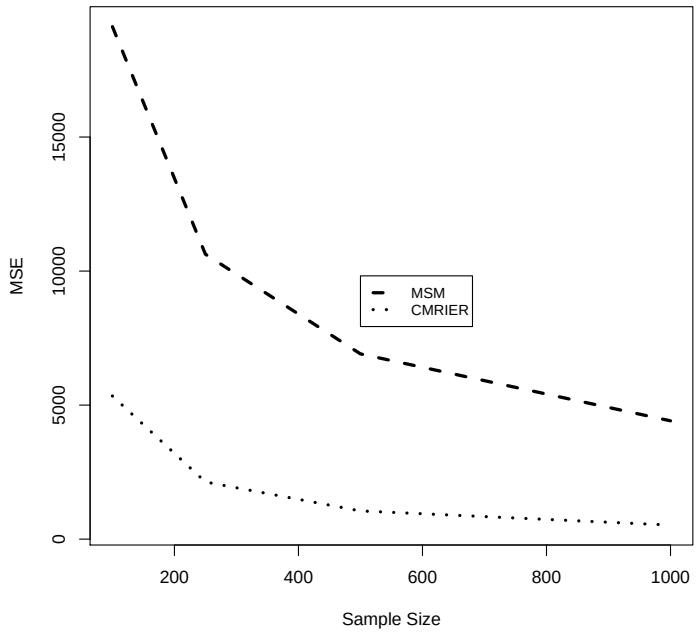
(b) β_1 and $\beta_1^{'}$

Figure 7: Simulation study with categorical exposure: IPTW estimator of MSM and CMRIER parameters, consistency

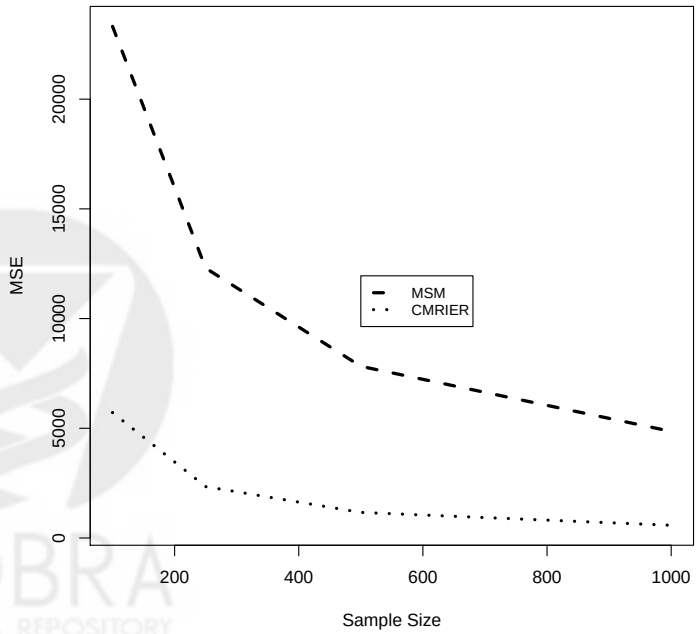


(c) β_2 and β_2'

Figure 7: Simulation Study with categorical exposure: IPTW estimator of MSM and CMRIER parameters, consistency, continued

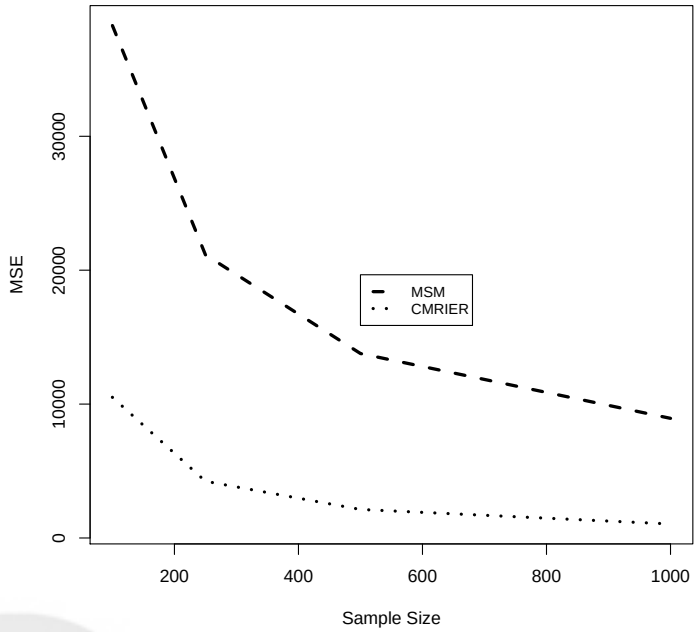


(a) β_0 and $\beta_0^{'}$



(b) β_1 and $\beta_1^{'}$

Figure 8: Simulation study with categorical exposure: IPTW estimator of MSM and CMRIER parameters, MSE



(c) β_2 and β_2'

Figure 8: Simulation study with categorical exposure: IPTW estimator of MSM and CMRIER parameters, MSE, continued

5.1.3 Simulation Summary

The simulation study demonstrates that in contrast to MSM parameters, CMRIER parameters can reliably be estimated in the presence of ETA violations. In both the binary and three-level categorical discretized exposure simulations, the behavior of the IPTW estimator of the CMRIER parameters largely outperformed the IPTW estimator of the MSM parameters with respect to both bias and MSE.

In addition to their usefulness in terms of mitigating estimation issues in the presence of ETA violations, the real data example provided in the next section demonstrates that they may be more sensible from a subject-matter perspective to draw causal inferences that can be fully supported by the information in the data.

5.2 Illustration with real data: air pollution study

The application of the CMRIER was motivated by the study briefly described throughout this paper, with the aim of estimating the extent to which reductions in ambient ozone concentrations were associated with measurable health benefits, specifically on the proportion of asthma-related hospital discharges. The fact that ozone is a continuous variable immediately raised concerns about possible ETA violations. Initial investigations did indeed show large practical violations of the ETA assumption. For this reason, the initial analysis relied on the G-computation estimator (Moore et al. (2008)). However, as mentioned previously, consistency of this estimator relies entirely on untestable model assumptions. This study demonstrated the need to aim for a parameter that was fully identifiable from the data by restricting the analysis to interventions for which the data carries information.

We apply the CMRIER methodology to this air pollution study that aimed to investigate the effect of reduction in ozone concentrations consequent to regulations propagated since 1980 by the California Air Resources Board in the Los Angeles Basin on the proportion of asthma-related hospital discharges (Moore et al. (2008)). The data consist of n=195 geographical grids, with quarterly measurements of confounders, quarterly average ambient ozone concentrations and quarterly proportions of asthma-related hospital discharges, where $W(t)$ is a multivariate vector of potential confounders measured at quarter t; $A(t)$ is the ozone concentration measured at quarter t and $Y(t)$ is the proportion of asthma-related hospital discharges measured in this same quarter t. The ozone concentration at quarter t, $A(t)$ was discretized to define $A^*(t)$, where $A^*(t) = 1$ if $A(t) > 90$, where 90 ppb was the California one-hour average ozone standard through the study years, and $A^*(t) = 0$ otherwise. We aim to investigate the the effect of $A^*(t-1)$ on $Y(t)$. Although the outcome at time t ($Y(t)$) and exposure at time $t-1$ ($A^*(t-1)$) are actually measured during the same quarter, we make the assumption that the exposure precedes the outcome.

For specifics as to the choice of the consideration of only a single quarter exposure as opposed to the whole history see Moore et al. (2008). Since only part of the exposure history is considered relevant, an HRMSM was applied to investigate the effect of interest (Neugebauer et al. (2007)). For the ease of presentation, we drop the t-notation and from this point on we refer only to A^* and Y. We modeled the expectation of counterfactual outcomes Y_{a^*} with the following HRMSM,

$$E(Y_{a^*}) = \beta_0 + \beta_1 a^*.$$

The coefficient β_1 can be interpreted as the population-level effect of the same quarter ozone level falling above versus below the 90ppb standard on the proportion of asthma-related hospital discharges during any given quarter (see Appendix A.1 for an explicit definition). As noted earlier, this parameter is only identifiable under the ETA assumption.

The Deletion/Subsitution/Addition (Sinisi and van der Laan (2004)) was applied to obtain the estimate $g_n(A \mid W)$ of the exposure mechanism from which $g_n(A^* \mid W)$ was obtained as described in Appendix A.1. Based on this estimate, the ETA is strongly violated, with 59% of the grids with probabilities of exposure less than 0.1 or greater than 0.9. Thus, it is apparent that it is not reasonable to assume that all grids could possibly be exposed to all levels of ozone at each point in time. This is an example where it is natural to reconsider the modeling approach to be used to answer the research question of interest. Based on subject matter and regulatory contexts, we consider modeling the population-level effect of ozone dichotomized above or below the standard on the proportion of asthma-related hospital discharges based on dynamic exposure interventions such that only those grids for which the intervention (ozone above or below the standard) is possible (in the sense of the realistic interventions) indeed receive the intervention.

To investigate this effect, we also applied the CMRIER-HRMSM,

$$E(Y_{d(a^*)(W)}) = \beta_0^{'} + \beta_1^{'} a^*,$$

where rule $d(a^*)(W)$ is defined in section 4.1, with $\alpha = 0.1$.

The IPTW estimator was used to estimate the parameters β and $\beta^{'}$ of the HRMSM and CMRIER-HRMSM respectively. The results are provided in Table 1. The point estimates of the HRMSM and CMRIER-HRMSM are drastically different, with negative effect and positive effect estimates, respectively.

The results from the original HRMSM analysis based on the continuous ozone variable estimated with the G-computation method resulted in an estimate of an increase of 1.44e-06 in the proportion of asthma-related hospital discharges for a one unit increase in ozone (Moore et al. (2008)).

Unlike results from the HRMSM analysis with the continuous ozone variable, the CMRIER results are not significant. Note that the HRMSM analysis was based on G-computation estimation which artificially relies on untestable parametric modeling assumptions to estimate HRMSM parameters when the ETA assumption is violated. Thus, in this ozone study, significant results from the G-computation analysis may be a consequence of the approach taken and not solely based on the information in the data. Moreover, the fact that the CMRIER analysis does not provide significant results may be due to the lack of power to detect an effect with inverse weighting estimation. A more efficient estimation approach like Targeted Maximum Likelihood estimation van der Laan and Rubin (2006) could improve the estimation precision for CMRIER parameters and may thus be implemented in the future in combination with CMRIER to provide a more definite conclusion regarding the effect of ozone in this study.

Table 1: Air pollution example: HRMSM and CMRIER-HRMSM results

	Estimate	Standard Error	p-value
HRMSM (β_1)	-3.89e-05	5.03e-05	0.44
CMRIER-HRMSM ($\beta_1^{'}$)	6.21e-06	7.52e-06	0.41

6 Discussion

In epidemiology, an increasing body of causal inference relies fundamentally on the concept of counterfactuals/potential outcomes. Causal modeling that explicitly relies on this concept is appealing for two reasons: 1) it allows the clear definition of the causal effect of interest; and 2) it reveals the critical assumptions on which causal inference from real data must rely: the no unmeasured confounders and the experimental treatment assignment assumptions. Marginal Structural Models were a break-through in causal inference because they allowed the development of a general modeling and estimation approach based on the counterfactual statistical framework with both point-treatment and longitudinal data. However, in practice, the validity of estimates based on their application has been hampered by frequent practical violations of the ETA assumption, on which accurate causal inference with MSM relies. This assumptions ensures that the data contain sufficient information to identify the causal effect of interest without requiring additional untestable parametric modeling assumptions to disentangle fully the effects of confounders from that of the exposure of interest. As described and illustrated in this article, several approaches have been proposed to draw reliable inference with MSM when the ETA assumption is violated. While these approaches can provide a satisfactory solution in practice in certain situations, they often fall short, as described in section 3. In addition, the ETA assumption is not unique to causal modeling with MSM; traditional regression models also rely on this assumption to draw reliable causal inference without making additional untestable parametric modeling assumptions.

The poor practical performance of MSM when the ETA assumption is violated should not be viewed as a drawback but an advantage of the approach. Indeed, unlike the traditional modeling approach, the MSM approach allows investigators to detect situations when the ETA assumption is violated or practically violated (Neugebauer and van der Laan (2005)). Detection of violations of the ETA assumption is essential to avoid drawing causal inferences that cannot be supported fully by the data. Moreover, it suggests that perhaps one should be aiming for a parameter that is identifiable with the data at hand.

Recent methodological work, based on the counterfactual framework, has resulted in the definition of a new class of causal models that generalize MSM: CMRIER. These models allow the investigation of causal effects that often may be more relevant for some subject-matter research, since these effects are defined based on dynamic exposure interventions which are actually possible for each of the units in the population. In addition, these models provide a general solution to the problem of causal inference when the ETA assumption is violated. The application of CMRIER for this purpose relies on the basic understanding of the interpretation of CMRIER parameters in order to properly define realistic exposure rules that will allow extraction of relevant information to address the original research question of interest. We provided a general methodology to define realistic exposure rules, but the investigator is not limited to the definition of causal effects based on such rules and, in fact, should consider carefully other alternative rules that could result in the definition of more interesting causal effects for their own research endeavors (Bembom and van der Laan (2007b)). In addition, while CMRIER provide an appealing solution to the problem of ETA violation in causal inference, their use in practice can be combined with other approaches that have been described in this article: exposure discretization, curtailed adjustment for

confounding or subgroup analysis. When the research question of interest involves the investigation of the effect of interventions on a history of exposure then CMRIER also can be used in combination with restriction of the exposure history as proposed with the HRMSM procedure referenced in this paper (Neugebauer et al. (2007)). Furthermore, note that under the often reasonable assumption that a causal effect represented by a CMRIER is smaller than the effect represented by an MSM, one can use a CMRIER to test that a causal effect represented by an MSM is null (no effect of static interventions) even when the ETA assumption does not allow identification of the MSM parameter with the observed data.

In this paper, the fitting of CMRIER focused on the IPTW estimation approach. Note however that the A-IPTW or targeted maximum likelihood approaches provide additional robustness to modeling assumptions and improved estimation efficiency/presicion. One appeal for a targeted maximum likelihood approach over an A-IPTW analysis is its ease of implementation (van der Laan and Rubin (2006)).

Finally, the general methodology described in this paper to define realistic exposure rules relies on an arbitrary choice for α which ensures that the probability of the exposure resulting from such exposure rules is not too close to 0. The choice for α is critical in two ways: 1) if chosen too small, the corresponding exposure rules are not realistic (finite-sample bias due to practical ETA violation remains a concern); and 2) if chosen too large, the causal effects defined based on these rules becomes less relevant and may be more difficult to detect. Therefore, the goal should be to select α as small as possible so that the CMRIER parameter is as close to the MSM parameter as possible while remaining identifiable from the data. Bembom and van der Laan (2008b) proposed the following approach to select and estimate the wished CMIER parameter: 1) Define a level of acceptable identifiability (as measured by bias, variance or both). Such a decision maps into a family of α levels, each of which ensures the definition of a CMRIER causal parameter that is deemed identifiable. The smallest α level in this family is chosen to define the CMRIER parameter of interest; 2) Estimate the CMIER parameter defined by the α level previously selected with the preferred method such as targeted maximum likelihood estimation involving cross-validation to select fine tuning parameters (van der Laan and Dudoit (2003); Dudoit and van der Laan (2005)). As suggested in Bembom and van der Laan (2008b), estimators of CMIER parameters defined by other α levels may be considered as candidate estimators of the actual CMIER parameter of interest. The estimator selected is chosen so that it minimizes the mean squared error associated with the CMIER parameter of interest.

APPENDIX

A Dichotomized exposure and effect definition

In this section, we briefly outline the framework on which the definitions of causal effects with dichotomized exposures is based. This framework enables the definition of causal effects based on not only static interventions (MSM) on the dichotomized exposure but also on dynamic interventions (CMRIER). Note that the framework detailed in this section is based on the extended counterfactual framework from which causal effect models for intention-to-treat were derived in van der Laan and Petersen (2007). In this section, we also describe

the procedure implemented in the simulation studies of this paper that can be used to derive the true value of causal effects with discretized exposures from the chosen data generating distribution along with a brief description of the maximum-likelihood and inverse-weighing estimator of the causal effects (for more details see van der Laan and Petersen (2007)). The results in this section can be generalized to discretization of exposures with more than two levels. Such a generalization was applied in the simulation studies from this paper when the continuous exposure was discretized in three categories.

A.1 Effect definition with MSM and dichotomized exposure

The observed data are $O = (W, A, Y)$ where A is a continuous exposure variable. We define the following sets of continuous treatment levels:

$$\mathcal{D}_0 \equiv \{a : a \leq \theta\} \text{ and } \mathcal{D}_1 \equiv \{a : a > \theta\},$$

where θ is the cut-off level used to dichotomize the continuous treatment into two categories. Similar to the approach taken to extend the counterfactual framework in van der Laan and Petersen (2007), we define the full data as:

$$X = (W, (A_{d(a^*)}, Y_{d(a^*)})_{a^* \in \mathcal{A}^*}),$$

where:

- $\mathcal{A}^* \equiv \{0, 1\}$ are the two possible levels for the dichotomized exposure.
- $A_{d(a^*)} \equiv AI(A \in \mathcal{D}_{a^*}) + A_{a^*} I(A \notin \mathcal{D}_{a^*})$.
- $g(A_{a^*} = a \mid W) = I(a \in \mathcal{D}_{a^*}) g(A = a \mid A \in \mathcal{D}_{a^*}, W)$.
- $Y_{d(a^*)} \equiv Y_{A_{d(a^*)}}$.

The observed data can then be viewed as a censored version of these full data with censoring variable $\Delta(a^*) \equiv I(A \in \mathcal{D}_{a^*})$:

$$O = (W, A, Y) = (W, (\Delta(a^*), \Delta(a^*)(A_{d(a^*)}, Y_{d(a^*)}))_{a^* \in \mathcal{A}^*}.$$

Based on this extended counterfactual framework, the following causal parameter $\beta = (\beta_0, \beta_1)$ can be defined:

$$E(Y_{d(a^*)}) = \beta_0 + \beta_1 a^* \text{ for } a^* \in \mathcal{A}^*.$$

This causal parameter represents the effect of a "low" versus "high" exposure to A where "low" and "high" are defined by the cut-off value θ. This parameter can be viewed as an MSM parameter defined by static interventions on the dichotomized exposure variable $A^* \equiv I(A > \theta)$ when the observed data are viewed as $O^* = (W, A^*, Y)$ and the full data are viewed as $X^* = (W, (Y_{a^*})_{a^* \in \mathcal{A}^*})$ where $Y_{a^*} \equiv Y_{d(a^*)}$:

$$E(Y_{a^*}) = \beta_0 + \beta_1 a^* \text{ for } a^* \in \mathcal{A}^*.$$

Likelihood-based estimation. Under the no unobserved confounder assumption, $A \perp X \mid W$ and the ETA assumption, $P(\min_{a^* \in \mathcal{A}^*} g(A \in \mathcal{D}_{a^*} \mid W) > 0) = 1$, the probability distribution of $(W, A_{d(a^*)}, Y_{d(a^*)})$ at (w, a, y) is identifiable with the observed data distribution through the following G-computation formula:

$$P(w, a, y) = P_W(w) g_{a^*}(a \mid w) P_{Y|A,W}(y \mid a, w),$$

where

$$\begin{aligned} g_{a^*}(a \mid w) &= g(A = a \mid W = w, A \in \mathcal{D}_{a^*}) \\ &= I(a \in \mathcal{D}_{a^*}) \frac{g(A = a \mid W = w)}{P(A \in \mathcal{D}_{a^*} \mid W = w)}. \end{aligned} \tag{1}$$

Given the marginal distribution of W, P_W, the conditional distribution of A given W, g_{a^*} (implied by g), and the conditional distribution of Y given A and W, $P_{Y|W,A}$, one can generate the counterfactuals $(W, A_{d(a^*)}, Y_{d(a^*)})$ as follows: 1) Generate W from P_W; 2) Generate $A_{d(a^*)}$ from distribution (2); 3) generate $Y_{d(a^*)}$ from $P_{Y|W,A}$. By applying this data generating experiment to an estimate of the data generating distribution, one obtains a large sample $(\hat{W}_b, \hat{A}_{d(a^*),b}, \hat{Y}_{d(a^*,b)})$, $b = 1, \ldots, B$ for all $a^* \in \mathcal{A}^*$ which yields a simulation-based estimate of the distribution of $(W, A_{d(a^*)}, Y_{d(a^*)})$. Such an estimate could now also be mapped into an estimate of β by least squares regression of the simulated $\hat{Y}_{d(a^*),b}$ on a^* according to the parametric model $m(a^* \mid \beta) = \beta_0 + \beta_1 a^*$. Note that in a simulation study, the observed data generating distribution is known and thus need not be estimated. This procedure applied with B very large will thus provide a good approximation of the true value of the causal parameter of interest in a simulation study. Note that to draw an observation from distribution (2), one can draw a from $g(A \mid W)$ until $a \in \mathcal{D}_{a^*}$. In our simulation study where the conditional distribution of A given W is Gaussian, the conditional distribution of A given W, $g_{a^*}(A \mid W)$, corresponds to a truncated normal distribution.

Inverse Weighting estimation. The following estimating function is unbiased for any non-null function h of $I(A > \theta)$:

$$\begin{aligned} D_h(O \mid \beta, g) &\equiv \frac{h(I(A > \theta))}{P(A \in \mathcal{D}_{I(A>\theta)} \mid W)} (Y - m(I(A > \theta) \mid \beta)) \\ &= I(A \leq \theta) \frac{h(I(A > \theta))}{P(A \leq \theta \mid W)} (Y - m(I(A > \theta) \mid \beta)) + \\ &\quad I(A > \theta) \frac{h(I(A > \theta))}{P(A > \theta \mid W)} (Y - m(I(A > \theta) \mid \beta)) \end{aligned}$$

If the observed data are viewed as O^* then this estimating function can be rewritten as:

$$D_h(O^* \mid \beta, g) = \frac{h(A^*)}{P(A^* \mid W)} (Y - m(A^* \mid \beta)).$$

If one chooses, $h(A^*) = \lambda(A^*) \frac{d}{d\beta} m(A^* \mid \beta)$, then the inverse probability of treatment weighted (IPTW) estimator of β can be obtained with the observed data O^* through a

weighted least squares regression of Y on A^* with the parametric model m and weights $\frac{\lambda(A^*)}{P(A^*|W)}$ where λ is any non-null function of A^*, e.g., $\lambda(A^*) = 1$ (unstabilized weighting) or $\lambda(A^*) = P(A^*)$ (stabilized weighting).

Note that in our simulation study where $g(A \mid W)$ is known (conditional Gaussian distribution), one can also derive $P(A^* \mid W)$ (the discretized exposure mechanism) as follows:

$$\begin{aligned} P(A^* = 1 \mid W) &= P(I(A > \theta) = 1 \mid W) \\ &= \int_{\theta}^{+\infty} g(a \mid W) da, \end{aligned}$$

and

$$\begin{aligned} P(A^* = 0 \mid W) &= P(I(A > \theta) = 0 \mid W) \\ &= \int_{-\infty}^{\theta} g(a \mid W) da. \end{aligned}$$

A.2 Effect definition with CMRIER and dichotomized exposure

We define the following sets of continuous treatment levels for $\alpha > 0$:

$$\mathcal{D}_w \equiv \{a^* : P(a^* \mid w) > \alpha\}$$

where α represents the probability level at which the dichotomized exposure is deemed realistic (e.g., $\alpha = 0.1$). Similar to the approach taken to extend the counterfactual framework above, we define the full data as:

$$X = (W, (A_{d(a^*)}, Y_{d(a^*)})_{a^* \in \mathcal{A}^*}),$$

where:

- $\mathcal{A}^* \equiv \{0, 1\}$ are the two possible levels for the dichotomized exposure.
- $A_{d(a^*)} \equiv I(a^* \in \mathcal{D}_W)\Big(AI(A \in \mathcal{D}_{a^*}) + A_{a^*} I(A \notin \mathcal{D}_{a^*})\Big) + I(a^* \notin \mathcal{D}_W)\Big(AI(A \in \mathcal{D}_{1-a^*}) + A_{1-a^*} I(A \notin \mathcal{D}_{1-a^*})\Big)$.
- $g(A_{a^*} = a \mid W) = I(a \in \mathcal{D}_{a^*}) g(A = a \mid A \in \mathcal{D}_{a^*}, W)$.
- $Y_{d(a^*)} \equiv Y_{A_{d(a^*)}}$.

The observed data can then be viewed as a censored version of these full data with censoring variable $\Delta(a^*) \equiv I(\ (a^* \in \mathcal{D}_W \ \&\ A \in \mathcal{D}_{a^*})$ or $(a^* \notin \mathcal{D}_W \ \&\ A \in \mathcal{D}_{1-a^*})\ )$:

$$O = (W, A, Y) = (W, (\Delta(a^*), \Delta(a^*)(A_{d(a^*)}, Y_{d(a^*)}))_{a^* \in \mathcal{A}^*}.$$

Based on this extended counterfactual framework, the following causal parameter $\beta = (\beta_0, \beta_1)$ can be defined:

$$E(Y_{d(a^*)}) = \beta_0 + \beta_1 a^* \text{ for } a^* \in \mathcal{A}^*.$$

Likelihood-based estimation. Under the no unobserved confounder assumption, $A \perp X \mid W$ and the ETA assumption, $P(\min_{a^* \in \mathcal{A}^*} P(I(A_{d(a^*)} > \theta) \in \mathcal{D}_W)) = 1$, the probability distribution of $(W, A_{d(a^*)}, Y_{d(a^*)})$ at (w, a, y) is identifiable with the observed data distribution through the following G-computation formula:

$$P(w, a, y) = P_W(w) g_{a^*}(a \mid w) P_{Y|A,W}(y \mid a, w),$$

where

$$\begin{aligned} g_{a^*}(a \mid w) &= g(A = a \mid W = w, A \in \mathcal{D}_{a^*})^{I(a^* \in \mathcal{D}_w)} g(A = a \mid W = w, A \in \mathcal{D}_{1-a^*})^{I(a^* \notin \mathcal{D}_w)} \\ g_{a^*}(a \mid w) &= \left(I(a \in \mathcal{D}_{a^*}) \frac{g(A = a \mid W = w)}{P(A \in \mathcal{D}_{a^*} \mid W = w)} \right)^{I(a^* \in \mathcal{D}_w)} * \\ &\quad \left(I(a \in \mathcal{D}_{1-a^*}) \frac{g(A = a \mid W = w)}{P(A \in \mathcal{D}_{1-a^*} \mid W = w)} \right)^{I(a^* \notin \mathcal{D}_w)}. \end{aligned} \tag{2}$$

Similar to the procedure described in the previous section with MSM, given the marginal distribution of W, P_W, the conditional distribution of A given W, g_{a^*} (implied by g), and the conditional distribution of Y given A and W, $P_{Y|W,A}$, one can generate the counterfactuals $(W, A_{d(a^*)}, Y_{d(a^*)})$ to obtain a large sample $(\hat{W}_b, \hat{A}_{d(a^*),b}, \hat{Y}_{d(a^*),b})$, $b = 1, \ldots, B$ for all $a^* \in \mathcal{A}^*$ which yields a simulation-based estimate of the distribution of $(W, A_{d(a^*)}, Y_{d(a^*)})$. Such an estimate could now also be mapped into an estimate of β by least squares regression of the simulated $\hat{Y}_{d(a^*),b}$ on a^* according to the parametric model $m(a^* \mid \beta) = \beta_0 + \beta_1 a^*$. Note that in a simulation study, the observed data generating distribution is known and thus need not be estimated. This procedure applied with B very large will thus provide a good approximation of the true value of the causal parameter of interest in a simulation study. Note that to draw an observation from distribution (2), one can draw a from $g(A \mid W)$ until i) $a \in \mathcal{D}_{a^*}$ if $a^* \in \mathcal{D}_W$ or ii) $a \in \mathcal{D}_{1-a^*}$ if $a^* \notin \mathcal{D}_W$.

Inverse Weighting estimation. The following estimating function is unbiased for any non-null function h of $I(A > \theta)$:

$$D_h(O \mid \beta, g) \equiv \sum_{a^* \in \mathcal{A}^*} \frac{\Delta(a^*)}{P(A \in \mathcal{D}_{I(A>\theta)} \mid W)} h(I(A > \theta))(Y - m(I(A > \theta) \mid \beta))$$

If the observed data are viewed as O^* then this estimating function can be rewritten as:

$$D_h(O^* \mid \beta, g) = \sum_{a^* \in \mathcal{A}^*} \frac{\Delta(a^*)}{P(A^* \mid W)} h(A^*)(Y - m(A^* \mid \beta)).$$

If one chooses, $h(A^*) = \lambda(A^*) \frac{d}{d\beta} m(A^* \mid \beta)$ (see previous section regarding choices for $\lambda(A^*)$), then the inverse probability of treatment weighted (IPTW) estimator of β can be obtained with the observed data O^* through weighted least squares regression of Y on A^* with the parametric model m where every observation in O^* is repeated for each $a^* \in \mathcal{A}^*$ and each such repeated observation is weighted separately with $\frac{\Delta(a^*)}{P(A^*|W)} \lambda(A^*)$ (see van der Laan and Petersen (2007) for more details).

References

Bembom, O. and van der Laan, M. (2007a). Analyzing sequentially randomized trials based on causal effect models for realistic individualized treatment rules. Technical Report 216, Division of Biostatistics, UC Berkeley.

Bembom, O. and van der Laan, M. (2007b). Estimating the effect of vigorous physical activity on mortality in the elderly based on realistic individualized treatment and intention-to-treat rules. Technical Report 217, Division of Biostatistics, UC Berkeley.

Bembom, O. and van der Laan, M. J. (2008a). Data-adaptive selection of the adjustment set in variable importance estimation. Technical Report 231, Division of Biostatistics, UC Berkeley.

Bembom, O. and van der Laan, M. J. (2008b). Data-adaptive selection of the truncation level for inverse-probability-of-treatment-weighted estimators. Technical Report 230, Division of Biostatistics, UC Berkeley.

Cole, S. R. and Hernan, M. A. (2008). Constructing Inverse Probability Weights for Marginal Structural Models. *Am. J. Epidemiol.*, page kwn164.

Dudoit, S. and van der Laan, M. J. (2005). Asymptotics of cross-validated risk estimation in estimator selection and performance assessment. *Statistical Methodology*, 2:131154.

Joffe, M., Knauss, J., Robinson, B., and Feldman, H. (2004). Lagged partially marginal structural models for controlling confounding. *Am. J. Epidemiol.*, 159((suppl)):S66.

Joffe, M., Santanna, J., and Feldman, H. (2001). Partially marginal structural models for causal inference. (abstract). *Am. J. Epidemiol.*, 153((suppl)):S261.

Moore, K., Neugebauer, R., Lurmann, F., Hall, J., Brajer, V., Alcorn, S., and Tager, I. (2008). Ambient ozone concentrations cause increased hospitalizations for asthma in children an 18-year study in southern california. *Environmental Health Perspectives*, 116(8):1063–1070.

Neugebauer, R. and van der Laan, M. J. (2003). Locally efficient estimation of nonparametric causal effects on mean outcomes in longitudinal studies. Technical Report 134, Division of Biostatistics, UC Berkeley.

Neugebauer, R. and van der Laan, M. J. (2005). Why prefer double robust estimators in causal inference? *Journal of Statistical Planning and Inference*, 129(1-2):405 – 426. IISA 2002 DeKalb Conference.

Neugebauer, R., van der Laan, M. J., Joffe, M., and Tager, I. (2007). Causal inference in longitudinal studies with history-restricted marginal structural models. *Electronic Journal of Statistics*, 1:119–154.

Neugebauer, R., v. d. L. M. J. (2006). Causal effects in longitudinal studies: Definition and maximum likelihood estimation. *Comput. Stat. Data Anal.*, 51(3):1664–1675.

Orellana L, Rotnitsky A, R. J. M. (2006). Generalized marginal structural models for estimating optimal treatment regimes. Technical report, Department of Biostatistics, Havard school of Public Health.

Robins, J. M. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math. Modelling*, 7(9-12):1393–1512. Mathematical models in medicine: diseases and epidemics, Part 2.

Robins, J. M. (1987). A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. *Journal of Chronic Disease*, 40:139S–161S.

Robins, J. M. (1998). Marginal structural models versus structural nested models as tools for causal inference. In *Symposium on Prospects for a Common Sense Theory of Causation*, AAAI Technical Report, Stanford, CA.

Robins, J. M. (2000). Robust estimation in sequentially ignorable missing data and causal inference models. In *Proceedings of the American Statistical Association*. American Statistical Association, Alexandria, VA.

Robins, J. M., Hernan, M. A., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560.

Robins, J. M., Orellana, L., and Rotnitzky, A. (2008). Estimation and extrapolation of optimal treatment and testing strategies. *Statistics in medicine*, 27(23):4678–4721.

Robins, J. M. and Rotnitzky, A. (2001). Comment on the Bickel and Kwon article, "Inference for semiparametric models: Some questions and an answer". *Statistica Sinica*, 11(4):920–936.

Sinisi, S. and van der Laan, M. J. (2004). The deletion/substitution/addition algorithm in loss function based estimation: Applications in genomics. *Statistical Applications in Genetics and Molecular Biology*, 3(1). Available from: `http://www.stat.berkeley.edu/~laan/Software/`.

van der Laan, M. J. (2006). Causal effect models for intention to treat and realistic individualized treatment rules. Technical Report 203, Division of Biostatistics, UC Berkeley.

van der Laan, M. J. and Dudoit, S. (2003). Unified cross-validation methodology for selection among estimators and a general cross-validated adaptive finite sieve estimator: Finite sample results, oracle inequalities, and examples. Technical Report 130, Division of Biostatistics, UC Berkeley.

van der Laan, M. J. and Petersen, M. (2007). Causal effect models for realistic individualized treatment and intention to treat rules. *The International Journal of Biostatistics*, 3(1).

van der Laan, M. J. and Robins, J. (2003). *Unified methods for censored longitudinal data and causality*. Springer, New York.

van der Laan, M. J. and Rubin, D. (2006). Targeted maximum likelihood learning. *The International Journal of Biostatistics*, 2(1):Article 11.

Wang, Y., Petersen, M., Bangsberg, D., and van der Laan, M. J. (2006). Data-adaptive selection of the adjustment set in variable importance estimation. Technical Report 211, Division of Biostatistics, UC Berkeley.

Yu, Z. and van der Laan, M. J. (2002). Construction of counterfactuals and the g-computation formula. Technical Report 122, Division of Biostatistics, UC Berkeley.